

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut National Polytechnique de Toulouse (Toulouse INP)

Discipline ou spécialité :

Informatique et Télécommunication

Présentée et soutenue par :

M. JULIEN DESVERGNES

le lundi 5 décembre 2022

Titre :

Apports multimodaux et interactifs pour améliorer la compréhension et la navigation de non-voyants utilisant un système de neuroprothèse visuelle

---

Ecole doctorale :

Mathématiques, Informatique, Télécommunications de Toulouse (MITT)

Unité de recherche :

Institut de Recherche en Informatique de Toulouse (IRIT)

Directeur(s) de Thèse :

M. CHRISTOPHE JOUFFRAIS

M. VINCENT CHARVILLAT

Rapporteurs :

M. FRANÇOIS CABESTAING, Université de Lille

MME EDWIGE PISSALOUX, UNIVERSITE DE ROUEN

Membre(s) du jury :

M. BENOÎT COTTEREAU, C.H.U. PURPAN, Président

M. AXEL CARLIER, TOULOUSE INP, Invité(e)

M. CHRISTOPHE JOUFFRAIS, TOULOUSE INP, Membre

MME EMILIE LOUP-ESCANDE, UNIVERSITE AMIENS, Membre

M. VINCENT CHARVILLAT, TOULOUSE INP, Membre

Titre : Apports multimodaux et interactifs pour améliorer la compréhension et la navigation de non-voyants utilisant un système de neuroprothèse visuelle

Résumé :

Le dernier rapport de l'OMS sur la non-voyance rapporte que 253 millions de personnes souffrent de déficience visuelle (36 millions d'entre elles sont aveugles et 217 millions ont une déficience visuelle). Pour aider les non-voyants dans leurs tâches quotidiennes, de nombreux dispositifs ont été conçus.

Les neuroprothèses visuelles sont des dispositifs apparus dans les années 60, qui permettent de rétablir une perception visuelle limitée chez les patients non-voyants implantés. Certaines de ces neuroprothèses, implantées dans la rétine ou dans le cortex visuel, comprennent un implant, un dispositif de calcul informatique et une caméra externe pour capturer la scène. Une perception visuelle appauvrie sous forme d'un ensemble de points blancs appelés phosphènes est alors restaurée grâce à une micro-stimulation de la rétine ou du cortex visuel via l'implant. Pendant les dix dernières années, plusieurs dispositifs ont été implantés sur des non-voyants et sont en phase d'essai clinique. Toutefois, la résolution (c'est-à-dire le nombre d'électrodes et leur densité) des implants actuels reste faible. Cette faible résolution, associée au nombre limité de couleurs différentes rendues par les implants, limite les informations qui peuvent être transférées et donc perçues.

Au cours de nos travaux, nous nous sommes intéressés à l'intérêt et à la manière de coupler des mécanismes interactifs et multimodaux avec une neuroprothèse visuelle, dans l'objectif d'améliorer les capacités de décisions des non-voyants. Il s'agit plus précisément de savoir dans quelles mesures l'ajout d'un système d'indices audio permet aux non-voyants de se déplacer plus facilement, et dans quelles mesures l'ajout d'un système de sélection interactif de rendus permet de désambiguïser la scène.

Dans ce mémoire, nous présentons les différentes technologies développées ces 30 dernières années ayant pour objectif de permettre à des non-voyants implantés de percevoir leur environnement en générant des cartes phosphéniques cohérentes. Nous abordons ensuite la question de la simulation de la vision prothétique, principalement due à la difficulté d'accès aux patients implantés. Ces simulateurs ont pour objectif de reproduire les conditions réelles de perception de phosphènes pour s'approcher le plus possible de la réalité. Ils permettent également de tester des méthodes de rendu originales. Nous présentons enfin deux études comparatives entre différents rendus, et nous montrons, d'une part, que l'ajout d'indices sonores couplés avec une restitution visuelle permet aux sujets de mieux comprendre la scène et de naviguer dans cette dernière, et, d'autre part, que donner aux non-voyants la possibilité d'alterner entre plusieurs modes de rendu en temps réel augmente significativement leur compréhension de l'environnement. En ce qui concerne l'ajout d'indices sonores, nous montrons que celui-ci permet aux sujets d'améliorer leurs

capacités de perception de la scène (détection et localisation d'obstacles), et procure aux sujets un sentiment de confiance. En ce qui concerne le système interactif, nous montrons qu'il permet aux sujets d'avoir un meilleur taux de bonnes réponses à une série de questions réparties sur trois catégories (objets, rues, portes et passages piétons). De plus, les sujets montrent un intérêt certain pour ce système interactif.

Title : Contributions in the communication of meaningful and useful information to blind people people using visual neuroprostheses

Abstract :
According to the WHO, 253 million people suffer from visual impairment : 36 million of them are blind and 217 million have a visual impairment between moderate and severe. In France, there are 1.7 million visually impaired people, including 207,000 blind and profoundly visually impaired people and about 900,000 moderately visually impaired people. People with visual impairments traditionally use a cane or a guide dog to navigate. But canes and guide dogs have a limited role in understanding the visual scene.

Visual neuroprostheses are devices that appeared in the 1960s, which allow the restoration of limited visual perception in blind implanted patients. Some of these neuroprostheses, localized in the retina or visual cortex, include an implant, a computing device and an external camera to capture the scene. An impoverished visual perception is then restored when the retina or visual cortex is microstimulated via the implant, leading to the appearance of white dots called phosphenes. Over the last ten years, several implants have been placed on blind people and are in clinical trials.

The difficulty of access to blind people (whether implanted or not) constraints us to use simulators. Their objective is to reproduce the real conditions of perception of phosphenes in order to get as close as possible to reality.

However, the resolution of the current implants (i.e. the number of electrodes and their spacing) that have passed the clinical trial phases remains low. This low resolution, associated with the limited number of different colors rendered by the implants, limits the information that can be transferred and thus perceived.

In this document we present the different technologies developed over the last 30 years with the objective of allowing blind people with implants to perceive their environment by generating coherent phosphenic maps. We have set up two comparative studies between different renderings and we have shown, on the one hand, that the addition of sound cues coupled with a visual rendering allows the subjects to better understand the scene and to navigate in it, and on the other hand, that giving the blind the possibility to alternate between several rendering modes in real time significantly increases their understanding of the environment.

Remerciements

Je n'aurais certainement pas pu conduire à bien ce projet de recherche aussi enrichissant que satisafaisant sans plusieurs personnes que j'aimerais remercier.

Pour commencer, je souhaite remercier Benoît Cottereau, d'avoir accepté d'être membre de mon jury de thèse et de le présider, mais également d'avoir été présent dès le début de ma thèse pour certains conseils avisés.

Ensuite, je voudrais remercier Edwige Pissaloux et François Cabestaing pour avoir accepté d'être rapporteurs de ce manuscrit de thèse. J'ai particulièrement apprécié leur patience, la qualité des retours et leurs commentaires constructifs.

Je souhaite de même remercier Émilie Loup-Escande pour avoir accepté de faire partie de mon jury et d'autant plus étant donné le délai de réponse très court que nous lui avions laissé.

Je remercie ensuite mon directeur de thèse Christophe Jouffrais, qui m'a permis de travailler sur ce sujet qui combine deux sujets qui me passionnent, l'informatique et la science du vivant. Christophe m'a toujours impressionné par la quantité de savoir qu'il possède et c'était toujours un plaisir d'échanger avec lui. Je remercie également Vincent Charvillat, mon co-directeur de thèse, qui a appuyé ma candidature et m'a fait confiance tout au long de la thèse. Je veux aussi remercier mon co-encadrant, Axel Carlier pour sa disponibilité, sa patience, sa bienveillance, son humour, son soutien et sa bonne humeur pendant ces trois années. Axel est la personne qui m'a le plus motivé et qui m'a contacté pour participer à ce projet. Si cette thèse a abouti c'est en grande partie grâce à lui.

Je souhaite remercier tous les permanents de l'équipe Reva. Je remercie Géraldine et Sylvie pour leur douceur et leur joie de vivre. C'était un plaisir d'enseigner à leur côté. Je remercie Jean-Denis pour ses attentions souvent bien placées et ses petites phrases qu'il m'arrive parfois de lui piquer. Je remercie Simone pour sa bonne humeur et les petites friandises qu'il ramenait de temps en temps, et qui ne faisait jamais très long feu.

Je remercie par ailleurs tous les autres membres de l'équipe que j'ai pu côtoyer : Bastien, Sonia, Matthieu, Damien, Paul, Thomas, Jean, Arthur, Thierry, Rémi, Marie, Travis, Baptiste, Déva, Hugo, Nicolas, Robin et Théo. Merci beaucoup pour tous ces moments partagés ensemble, que ce soit autour d'un repas, dans un karaoké ou encore en parties endiablées de coinche et autre jeux de sociétés. Je remercie particulièrement Thomas pour l'aide qu'il m'a apportée lors de la soutenance et Thierry qui m'a permis d'aiguiser mon sens du sarcasme à un niveau stratosphérique.

Je remercie particulièrement Sylvie et André-Luc pour toute la gestion administrative liée à ma thèse, surtout à l'approche de ma soutenance.

Enfin, je remercie mes parents, ma sœur et mon frère pour m'avoir soutenu et motivé

pendant ces trois ans. Pour finir, je remercie profondément Sofia qui est celle qui m'a toujours motivé et soutenu, qui a été et est toujours le pillier de ma vie.

Liste des publications

— Desvergnes, J., Carlier, A., Charvillat, V., Jouffrais, C. (2022, June). Une neuroprothèse visuelle interactive pour mieux percevoir l'environnement. In 12ème Conférence Handicap 2022 «Humaines et artificielles, les intelligences au service du handicap»,

— Desvergnes, J., Carlier, A., Ooi, W. T., Charvillat, V., Jouffrais, C. (2022, July). Does switching between different renderings allow blind people with visual neuroprostheses to better perceive the environment ?. In ICCHP AAATE 2022 (Joint International Conference on Digital Inclusion, Assistive Technology Accessibility).

Table des matières

Introduction xxvii

1 Système visuel humain, déficience visuelle et état de l'art sur la vision prothétique 1

1.1 Système visuel humain . . . 2
1.1.1 L'œil . . . 2
1.1.2 Le nerf optique . . . 6
1.1.3 Le cortex visuel . . . 6
1.2 Déficience visuelle et caractérisation . . . 8
1.2.1 Classification . . . 9
1.2.2 Dégénérescence maculaire liée à l'âge (DMLA) . . . 9
1.2.3 Rétinite pigmentaire (RP) . . . 10
1.3 Neuroprothèses visuelles . . . 11
1.3.1 Neuroprothèses rétiniennes . . . 11
1.3.2 Neuroprothèses corticales . . . 23
1.3.3 Neuroprothèses dans le corps géniculé latéral . . . 26
1.3.4 Neuroprothèses sur le nerf optique . . . 26
1.3.5 Tableau récapitulatif des neuroprothèses présentées . . . 27
1.4 Simulateurs de vision prothétique . . . 28
1.4.1 Problèmes liés à l'utilisation de neuroprothèses visuelles . . . 29
1.4.2 Quelques simulateurs de la littérature . . . 32
1.4.3 Tableau récapitulatif des simulateurs recensés pour des tâches de navigation ou de manipulation/reconnaissance d'objets . . . 42
1.5 Apprentissage profond et réseaux de neurones . . . 43
1.5.1 Historique . . . 43
1.5.2 Nos besoins en termes de réseaux de neurones . . . 46

2 Simulateur de vision prothétique 53

2.1 Unity 3D . . . 54
2.1.1 Présentation générale . . . 54
2.1.2 Aspect technique . . . 55

2.2 Construction de la représentation phosphénique . . . 58
2.2.1 Résolution de l'implant . . . 59
2.2.2 Taux d'électrodes défectueuses . . . 59
2.2.3 Géométrie variable des phosphènes . . . 62
2.2.4 Espacement inter-phosphènes . . . 62
2.2.5 Nombre de niveaux de gris variable . . . 63
2.2.6 Flou gaussien . . . 64
2.2.7 Tous les défauts en même temps . . . 64
2.3 Extraction des caractéristiques et rendu . . . 64
2.3.1 "Scoreboard" . . . 66
2.3.2 Contour brut . . . 68
2.3.3 Segmentation sémantique et extraction d'objets . . . 68
2.3.4 Contour amélioré . . . 69
2.3.5 Structure seule . . . 70
2.3.6 Méthode combinée . . . 71
2.4 Fonctionnement du simulateur . . . 72
2.4.1 Système de caméra . . . 72
2.4.2 Sélection des paramètres . . . 74
2.4.3 Comparaison des méthodes de restitutions . . . 74

3 Des indices audio pour améliorer la navigation et la compréhension avec une neuroprothèse 79

3.1 Utilisation de systèmes sonores dans l'aide à la navigation et la compréhension 81
3.1.1 Systèmes basés lunettes . . . 81
3.1.2 Systèmes basés canne . . . 84
3.2 Méthode et Matériel . . . 87
3.2.1 Simulateur de vision prothétique . . . 87
3.2.2 Rendus phosphéniques . . . 89
3.2.3 Hypothèses . . . 89
3.2.4 Sujets et conditions expérimentales . . . 90
3.3 Protocole . . . 91
3.3.1 Formulaire et présentation . . . 91
3.3.2 Familiarisation avec le dispositif expérimental . . . 91
3.3.3 Bloc d'expérience . . . 93
3.3.4 Évaluation finale . . . 95
3.4 Résultats . . . 97
3.4.1 Hypothèse 1 : Amélioration du comportement de navigation des sujets 97
3.4.2 Hypothèse 2 : Amélioration des capacités d'analyse et de perception du sujet . . . 97

3.4.3 Hypothèse 3 : Préférences de l'utilisateur . . . 100
3.5 Analyse des résultats et discussion . . . 102
3.5.1 Comportement des sujets . . . 102
3.5.2 Capacités d'analyse et de perception du sujet . . . 102
3.5.3 Satisfaction et classement . . . 102
3.5.4 Concernant les différents type de rendus . . . 102
3.5.5 Concernant la simulation . . . 103
3.5.6 Concernant les retours audio . . . 103

4 Une neuroprothèse visuelle interactive pour mieux percevoir l'environnement 107
4.1 Matériel et méthodes . . . 109
4.1.1 Simulateur de vision prothétique . . . 109
4.1.2 Rendus phosphéniques . . . 114
4.1.3 Hypothèse . . . 114
4.1.4 Sujets . . . 115
4.1.5 Conditions expérimentales . . . 115
4.2 Protocole . . . 116
4.2.1 Protocole expérimental . . . 116
4.2.2 Variables analysées . . . 117
4.3 Résultats . . . 119
4.3.1 Comparaison des trois rendus utilisés . . . 119
4.3.2 Jugement subjectif . . . 123
4.3.3 Classement final . . . 124
4.3.4 Analyse détaillé du comportement observé pour le rendu Switch . . . 124
4.4 Analyse et discussion des résultats . . . 126
4.5 Limitations de l'étude et perspectives . . . 129

Conclusion et perspectives 133
1 Amélioration du rendu phosphénique . . . 134
1.1 Phénomène d'évanescence . . . 134
1.2 Phénomène de persistance . . . 134
1.3 Géométrie des phosphènes . . . 136
1.4 Apparition de phosphènes multiples . . . 137
2 Amélioration des conditions expérimentales . . . 138
2.1 Module de suivi du regard . . . 138
2.2 Utilisation d'un casque de réalité virtuelle - entre simulation et environnement réel . . . 139
2.3 Intégration des réseaux de neurones et validation des résultats en cas réel . . . 141

3 Autres expériences réalisables 142
3.1 Switch + Audio 142
3.2 Encombrement 142
3.3 Apprentissage par renforcement 142

Annexe **145**
MeCue - Etude 1 145
Questions Likert - Etude 2 147

Bibliographie **149**

Résumés **163**
Popularized abstract 163
Résumé vulgarisé 163
Abstract 164
Résumé 164

Table des figures

1.1 Coupe de l'œil humain . . . 3
1.2 Densité des photorécepteurs en fonction de l'excentricité sur la rétine de l'oeil gauche . . . 4
1.3 Le montage de l'œil, un système à l'envers . . . 5
1.4 Le montage de l'œil transformé en mode "appareil photo" . . . 5
1.5 Nerf et chiasma optique . . . 6
1.6 Cartographie rétinotopique, deux hémisphères . . . 7
1.7 Cartographie rétinotopique, hémisphère gauche . . . 7
1.8 Coupe sagittale du cortex, face interne . . . 8
1.9 Focus sur les zones V1, V2 et V3 du cortex . . . 8
1.10 Grille d'évaluation de l'acuité visuelle . . . 10
1.11 Dégénérescence maculaire liée à l'âge (DMLA) et Rétinite pigmentaire (RP) 11
1.12 Sites d'implantation pour des implants rétiniens . . . 12
1.13 Neuroprothèse Argus II . . . 13
1.14 Résultats d'expériences de perception et navigation dans le cas de l'Argus II 14
1.15 Neuroprothèse IRIS II . . . 16
1.16 Résultats d'expériences de perception dans le cas de l'IRIS II . . . 17
1.17 Neuroprothèse EPI-RET 3 . . . 17
1.18 Neuroprothèse BRIP - Autour de l'œil . . . 18
1.19 Neuroprothèse BRIP - Composants . . . 19
1.20 Neuroprothèse ALPHA IMS - Composants sur le système visuel humain . . . 20
1.21 Neuroprothèse ALPHA IMS - Focus sur la matrice d'électrodes . . . 20
1.22 Neuroprothèse PRIMA . . . 21
1.23 Neuroprothèse BVA . . . 23
1.24 Neuroprothèse GENNARIS - Composants extrenes . . . 24
1.25 Neuroprothèse GENNARIS - Composants internes . . . 24
1.26 Neuroprothèse ORION . . . 25
1.27 Neuroprothèse C-SIGHT . . . 27
1.28 Non conservation de la cartographie rétinotopique . . . 30
1.29 Phénomène d'évanescence . . . 30
1.30 Résolution des implants limités . . . 31

1.31 Simulateur de Cha et al. - Tâche d'ouverture de porte 33
1.32 Simulateur de Cha et al. - Différentes résolutions 33
1.33 Simulateur de Cha et al. - Tâche de navigation 34
1.34 Simulateur de Dagnelie et al. - Tâche de navigation en conditions réelles . . 35
1.35 Simulateur de Dagnelie et al. - Tâche de navigation dans environnement virtuel simulé 35
1.36 Simulateur de Denis et al. - Rendus utilisés dans une tâche de préhension d'objets 36
1.37 Simulateur de Mc Carthy et al. - Rendus phosphéniques 37
1.38 Simulateur de Mc Carthy et al. - Sujet en expérience de navigation 38
1.39 Simulateur de Zapf et al., 2014 - Rendus phosphéniques 38
1.40 Simulateur de Zapf et al., 2014 - Tâche de navigation 39
1.41 Simulateur de Zapf et al., 2015 - Vision tubulaire 39
1.42 Simulateur de Zapf et al., 2015 - Tâche de navigation 40
1.43 Simulateur de Vergnieux et al., 2017 - Rendus phosphéniques 41
1.44 Simulateur de Vergnieux et al., 2017 - Plans tracés par les sujets 41
1.45 Simulateur de Sanchez-Garcia et al., 2019 - Rendus phosphéniques 42
1.46 Simulateur de Sanchez-Garcia et al., 2019 - Rendu combiné 42
1.47 Fonctionnement du neurone formel 45
1.48 Réseau convolutionnel LeNet5 45
1.49 Segmentation sémantique de scène urbaine 47
1.50 Segmentation sémantique sur des images d'extérieur avec une base ResNet-50 48
1.51 Segmentation sémantique réalisée par le réseau DeepLab V3+ 48
1.52 Extraction de structure utilisant des réseaux de neurones 49
1.53 Extraction de structure - Panoroom 49
1.54 Extraction de structure - LayoutNet 50

2.1 Illustration gameplay - Furi 55
2.2 Illustration gameplay - Haven 55
2.3 Illustration gameplay - The Heretic 56
2.4 Scène d'intérieur en Unity 3D 56
2.5 Illustration gameplay - Escape from Tarkov 57
2.6 Illustration gameplay - Kerbal Space Program 58
2.7 Construction de la représentation phosphénique - Résolution de l'implant . 60
2.8 Construction de la représentation phosphénique - Taux d'électrodes défectueuses 61
2.9 Construction de la représentation phosphénique - Géométrie variable des phosphènes 62
2.10 Construction de la représentation phosphénique - Espacement inter-phosphènes 63

2.11 Construction de la représentation phosphénique - Nombre de niveaux de gris variable . . . 65
2.12 Construction de la représentation phosphénique - Flou gaussien . . . 66
2.13 Construction de la représentation phosphénique - Tous les défauts en même temps . . . 67
2.14 Rendu scoreboard - Extraction des caractéristiques et rendu . . . 67
2.15 Rendu contour brut - Extraction des caractéristiques et rendu . . . 68
2.16 Rendu segmentation sémantique . . . 69
2.17 Rendu segmentation sémantique d'objets . . . 70
2.18 Rendu contour amélioré . . . 70
2.19 Rendu structure seule . . . 71
2.20 Rendu combiné . . . 72
2.21 Système de caméras - Acquisition . . . 73
2.22 Système de caméras - Rendu . . . 73
2.23 Réglages des paramètres de la restitution phosphénique . . . 74
2.24 Comparaison des méthodes de restitutions . . . 75

3.1 Dispositif d'aide sonore mis en place par l'équipe de Sadi et al. en 2014 . . . 82
3.2 Dispositif d'aide sonore mis en place par l'équipe de Sadi et al. - Caractéristiques des retours sonores . . . 82
3.3 Dispositif d'aide sonore mis en place par l'équipe de Kassim et al. en 2016 . . . 83
3.4 Dispositif d'aide par vibration mis en place par l'équipe de Kassim et al. en 2016 . . . 83
3.5 Dispositif d'aide sonore mis en place par l'équipe de Yi et al. en 2016 . . . 84
3.6 Dispositif d'aide sonore mis en place par l'équipe de Stoll et al. en 2015 . . . 85
3.7 Prototype de canne proposé par Fan et al. . . . 86
3.8 Exemple de scène utilisé dans notre première étude . . . 88
3.9 Scènes modélisées dans l'étude 1 . . . 89
3.10 Interface de l'étude 1 - Sélection des objets perçus . . . 94
3.11 Interface de l'étude 1 - Sélection du plan de la scène . . . 95
3.12 Interface de l'étude 1 - Placement des objets perçus dans la scène . . . 95
3.13 Interface de l'étude 1 - Questionnaire de satisfaction MeCue . . . 96
3.14 Interface de l'étude 1 - Classement des rendus . . . 96
3.15 Résultats de l'étude 1 - Temps de parcours par condition . . . 98
3.16 Résultats de l'étude 1 - Nombre de collisions par condition . . . 98
3.17 Résultats de l'étude 1 - Nombre d'objets reconnus . . . 99
3.18 Résultats de l'étude 1 - Nombre d'objets bien localisés . . . 100
3.19 Résultats de l'étude 1 - MeCue . . . 101
3.20 Résultats de l'étude 1 - Classement final . . . 101

4.1 Rendu basé sur l'approche proposée par Sanchez-Garcia et al. en 2020 . . . 108
4.2 Scène utilisé dans l'étude 2, basée sur le travail de Paul PJ Maget 110
4.3 Interface du mode Switch 115
4.4 Interface de l'étude 2 - Réponse aux questions 120
4.5 Interface de l'étude 2 - Questionnaire subjectif 120
4.6 Interface de l'étude 2 - Classement final 120
4.7 Résultats de l'étude 2 - Bonnes réponse aux questions 121
4.8 Résultats de l'étude 2 - Réponse "Je ne sais pas" 122
4.9 Résultats de l'étude 2 - Temps de réponse 123
4.10 Résultats de l'étude 2 - Jugement subjectif 124
4.11 Résultats de l'étude 2 - Classement final 125
4.12 Résultats de l'étude 2 - Temps d'utilisation des trois modes de rendu du Switch 125
4.13 Résultats de l'étude 2 - Nombre d'utilisation des trois modes de rendu du Switch 126
4.14 Résultats de l'étude 2 - Dernier rendu utilisé dans le mode Switch 127
4.15 Résultats de l'étude 2 - Évolution du temps de réponse en fonction de la condition utilisée 128
4.1 Perspective - Modélisation de l'évanescence phosphénique 135
4.2 Perspective - Modélisation de la persistance phosphénique 136
4.3 Perspective - Module de suivi du regard 139
4.4 Casque de réalité virtuelle Vive Pro Eye 140
4.5 Casque de réalité virtuelle Vive Flow 140
4.6 Perspective - Apprentissage par renforcement 143

Liste des tableaux

1.1 Tableau récapitulatif des implants 28
1.2 Tableau récapitulatif des simulateurs de vision prothétique 44

3.1 Placement des objets sur un plan réalisé par des sujets 92
3.2 Déplacement des sujets dans la scène pendant la tâche de navigation 93

4.1 Rendus utillisés dans l'étude 2 - Catégorie "objets" 111
4.2 Rendus utillisés dans l'étude 2 - Catégorie "rues" 112
4.3 Rendus utillisés dans l'étude 2 - Catégorie "portes et passages piétons" . . . 113
4.4 Questions posées dans l'étude 2 - Catégorie "objets" 117
4.5 Questions posées dans l'étude 2 - Catégorie "rues" 118
4.6 Questions posées dans l'étude 2 - Catégorie "portes et passages piétons" . . 119
4.7 Construction de la représentation phosphénique - Géométrie des phosphènes 137

Introduction

Selon l'OMS [Bourne et al., 2017], 253 millions de personnes souffrent de déficience visuelle : 36 millions d'entre elles sont aveugles et 217 millions ont une déficience visuelle entre modérée et sévère. En France, on compte, 1,7 million de personnes déficientes visuelles dont 207 000 aveugles et malvoyants profonds et environ 900 000 malvoyants. Les deux principales maladies affectant la vision sont la dégénérescence maculaire liée à l'âge (DMLA) et la rétinite pigmentaire (RP), qui touchent respectivement les personnes vieillissantes et les enfants. Traditionnellement, les personnes atteintes d'une déficience visuelle utilisent une canne ou un chien-guide pour se déplacer. Mais les cannes et les chiens-guides ont un rôle limité dans la compréhension de la scène visuelle. Avec les progrès de l'informatique et de la vision par ordinateur, de nombreux systèmes ont vu le jour. Ils sont le produit de recherches qui augmente les perceptions sensorielles des personnes visuellement handicapées en combinant les informations des capteurs visuels et d'autres données auxiliaires (comme la localisation GPS), en traitant les données visuelles à l'aide de techniques d'intelligence artificielle (comme la reconnaissance des visages et des émotions) et en présentant les informations à l'utilisateur à l'aide d'interfaces audio, tactiles [Conradie et al., 2014], ou les deux.

C'est dans ce contexte que dans les années 60, les neuroprothèses visuelles sont apparues [Wilson and Dorman, 2008a, Wilson and Dorman, 2008b]. Au cours de la dernière décennie, elles se sont placées comme une technique prometteuse pour restaurer partiellement la vision chez les personnes atteintes de déficience visuelle [Fernandez, 2018]. Il existe à ce jour deux types de neuroprothèses. Le premier est un dispositif implanté dans la rétine et basé sur des photorécepteurs, appelé MPDA pour *microphotodiode-array*. Les photons frappent ces derniers et stimulent alors les neurones sous-jacents (ALPHA IMS [Stingl et al., 2015]). Dans le second dispositif, une caméra externe capture la scène et envoie des signaux vers un implant qui stimule la rétine ou le cortex pour restaurer des perceptions visuelles. Les implants rétiniens peuvent être épirétinien (sur la rétine), sous-rétinien (juste en dessous de la rétine) ou encore supra-choroïdien (entre la choroïde et la sclérotique). Les tâches lumineuses engendrées dans le champ de vision suite à la stimulation d'une électrode s'appellent des phosphènes.

Depuis les dix dernières années, plusieurs implants ont été placés sur des non-voyants et sont en phase d'essais cliniques. Les neuroprothèses utilisant ce genre d'implants ne sont

toutefois pas totalement fonctionnelles en raison d'une large variété de problèmes, dont une résolution limitée, c'est-à-dire du faible nombre d'électrodes utilisées (le système PRIMA en comptait 378 lors d'un essai clinique en 2020 [Palanker et al., 2020]), qui ne permet pas de transmettre suffisamment d'informations pour rendre compte d'une scène visuelle. Une forte limitation des niveaux d'intensité lumineuse perceptibles rend la tâche encore plus complexe. De plus, le bruit (imperfection dans le placement de la matrice d'électrodes, dysfonctionnement des électrodes) dans l'affichage des phosphènes peut également poser des difficultés de perception.

Les travaux présentés dans ce mémoire tentent de répondre à la question suivante :

Dans quelle mesure est-il possible de communiquer des informations significatives et utiles à des sujets implantés utilisant un système de neuroprothèse visuelle par l'intermédiaire de dispositifs multimodaux et interactifs ?

Par la suite, nous prendrons pour référence le manuel d'intervention en orientation et mobilité proposé par Agathe Ratelle et Julie-Anne Couturier [Ratelle, 2019] qui recense une liste de besoins émanant directement de non-voyants. Ces besoins sont classés en deux catégories : les besoins en situation intérieurs et les besoins en situation extérieurs. Certains des besoins identifiés sont communs aux deux situations, comme par exemple pouvoir connaître la nature d'un objet qui fait face (banc public, chaise, …), mais certains besoins sont spécifiques à certains environnements : pouvoir attraper un objet sur une table ou pouvoir franchir un passage piéton en toute sécurité.

La question de la sélection des informations utiles se pose alors. Les informations à communiquer dans un contexte de navigation en scène d'extérieur ne sont pas du tout les mêmes que dans le cas d'une tâche de préhension d'objet en intérieur (notion de risque différente, taille des éléments différents, …). Il est donc nécessaire d'avoir à notre disposition des outils capables d'extraire des informations précises de la scène visuelle. Les réseaux de neurones et l'apprentissage profond permettent de répondre à cette problématique. Grâce à ces outils, il est possible d'extraire en temps réel des informations (emplacement des objets dans la scène, taille de ces objets, jointures entre trottoir et route). Ce sont ces informations qui vont nous permettre de sélectionner quelles électrodes il faut stimuler pour construire une représentation phosphénique qui a du sens dans le champ de vision du non-voyant implanté.

Malheureusement, il est extrêmement difficile d'avoir accès à des non-voyants implantés afin de pouvoir réaliser des expériences pour plusieurs raisons. Les deux principales sont les suivantes : - il y a très peu de patients implantés en France (environ 1000) et même à travers le monde ; - les patients implantés sont protégés par les sociétés de développement d'implants et de neuroprothèses pour des raisons de sécurité (n'importe qui ne doit pas pouvoir mettre n'importe quel programme sur les neuroprothèses). C'est pourquoi nous

sommes obligés de réaliser des expérimentations en situation contrôlées et donc d'utiliser des simulateurs de vision prothétique. La majorité des études menées dans ce domaine reposent sur des simulateurs de vision prothétique avec une capacité de représentation plus ou moins proche des contraintes réelles. De nombreuses études modélisent les implants de manière trop éloignée de la réalité, par exemple en utilisant des résolutions d'implants supérieures à celles implantables, limitant ainsi le caractère transmissible aux non-voyants. En pratique, ces expériences sont conduites grâce à des sujets voyants placés dans le simulateur. Un autre intérêt qui apparaît lors de la simulation de vision prothétique est que cela permet de tester des méthodes qui ne sont pas possibles aujourd'hui sur les implants existants. Il est par exemple possible d'observer le nombre minimal d'électrodes nécessaire pour effectuer une tâche précise, ou évaluer un score de réussite sur une tâche dépendamment de l'allure des phosphènes générés.

Nous essaierons de répondre à la problématique posée en évaluant deux systèmes d'aide différents. Le premier système est multimodal, et consiste à coupler une restitution visuelle prothétique à un mécanisme d'indices audio qui permet de donner des indications sémantiques sur la composition de la scène aux sujets. Nous mettons en place une étude qui tente de répondre aux besoins liés à la compréhension et à la navigation dans une scène intérieure. Le second système est interactif, et permet de donner le contrôle aux sujets sur la méthode de restitution visuelle utilisée par l'implant en temps réel. Ils peuvent ainsi changer de méthode selon leur volonté. Ce système tente de répondre aux besoins exprimés par les non-voyants en ce qui concerne la compréhension de scènes extérieures et plus particulièrement la compréhension des jonctions et intersections de routes, la détection de passages piétons et la détection de portes d'entrées.

La suite de ce mémoire s'organise de la manière suivante. Le chapitre 1 est consacré à la présentation du système visuel humain, de la caractérisation de la déficience visuelle, des différents types de neuroprothèses visuelles, des simulateurs de vision prothétiques existant et de l'intérêt de l'apprentissage profond dans notre contexte. Le chapitre 2 présente le simulateur de vision prothétique développé au cours des trois dernières années par notre équipe. Suivent ensuite les chapitres 3 et 4 qui présentent respectivement nos travaux liés à l'ajout d'indices sonores pour augmenter la compréhension de scènes d'intérieur, et ceux liés à l'intégration d'un module de sélection interactif des rendus. Nous concluons et présentons quelques perspectives dans la dernière partie.

Chapitre 1

Système visuel humain, déficience visuelle et état de l'art sur la vision prothétique

Sommaire

1.1 Système visuel humain . . . 2
1.1.1 L'œil . . . 2
1.1.2 Le nerf optique . . . 6
1.1.3 Le cortex visuel . . . 6
1.2 Déficience visuelle et caractérisation . . . 8
1.2.1 Classification . . . 9
1.2.2 Dégénérescence maculaire liée à l'âge (DMLA) . . . 9
1.2.3 Rétinite pigmentaire (RP) . . . 10
1.3 Neuroprothèses visuelles . . . 11
1.3.1 Neuroprothèses rétiniennes . . . 11
1.3.2 Neuroprothèses corticales . . . 23
1.3.3 Neuroprothèses dans le corps géniculé latéral . . . 26
1.3.4 Neuroprothèses sur le nerf optique . . . 26
1.3.5 Tableau récapitulatif des neuroprothèses présentées . . . 27
1.4 Simulateurs de vision prothétique . . . 28
1.4.1 Problèmes liés à l'utilisation de neuroprothèses visuelles . . . 29
1.4.2 Quelques simulateurs de la littérature . . . 32
1.4.3 Tableau récapitulatif des simulateurs recensés pour des tâches de navigation ou de manipulation/reconnaissance d'objets . . . 42
1.5 Apprentissage profond et réseaux de neurones . . . 43
1.5.1 Historique . . . 43
1.5.2 Nos besoins en termes de réseaux de neurones . . . 46

L'intérêt porté aux systèmes permettant de restituer ou de pallier la vision n'a cessé de croître pendant les trente dernières années. Que ce soit en ce qui concerne la création de neuroprothèses ou de simulateurs de vision prothétique, les apports de ces recherches ont permis l'implantation sur des non-voyants de plusieurs dispositifs de restitution visuelle.

Ce chapitre a pour but de présenter dans un premier temps le système visuel humain (1.1) ainsi que les causes de la déficience visuelle (1.2). Ensuite, nous dressons une liste des différentes neuroprothèses développées (1.3), puis nous abordons les outils de simulation de vision, qui permettent de valider un processus de restitution visuelle (1.4). Enfin, nous présentons un outil informatique qui permet de construire des rendus sophistiqués : les réseaux de neurones (1.5).

1.1 Système visuel humain

1.1.1 L'œil

Chez l'être humain, la vue est la source d'information principale sur l'environnement extérieur. L'œil, (dont une coupe est présentée sur la FIGURE 1.1 [1]) est un des organes principaux de la vision, avec le nerf optique et le cerveau. Il a pour fonction de capter la lumière et de la transformer en impulsions électriques qui seront acheminées au cortex par l'intermédiaire du nerf optique. Ces impulsions sont traitées par les cellules localisées entre les photorécepteurs et le cortex visuel, qui traduit les informations et nous les renvoie sous forme d'images, nous permettant ainsi d'interpréter notre environnement.

La cornée

La cornée est le premier milieu transparent rencontré par les rayons lumineux. Sa forme aide à la transmission et à la convergence des rayons lumineux sur la fovéa. On retrouve de nombreuses pathologies qui peuvent atteindre la cornée : kératite, abcès, érosion, ulcère, kératocône. La cornée étant largement innervée, toute pathologie aiguë de la cornée est douloureuse. Les maladies qui altèrent la transparence cornéenne entraînent systématiquement une baisse d'acuité visuelle.

Humeur aqueuse

L'humeur aqueuse est le premier milieu liquide traversé par les rayons lumineux. Elle est située entre l'iris et la cornée. Ce liquide contient également du glucose et du dioxygène qui sert à alimenter le cristallin et la cornée. L'humeur aqueuse se renouvelle de manière continue, ce qui lui permet d'éliminer les déchets des cellules proches.

1. https://www.alloprof.qc.ca/fr/eleves/bv/sciences/la-vue-et-l-oeil-s1288

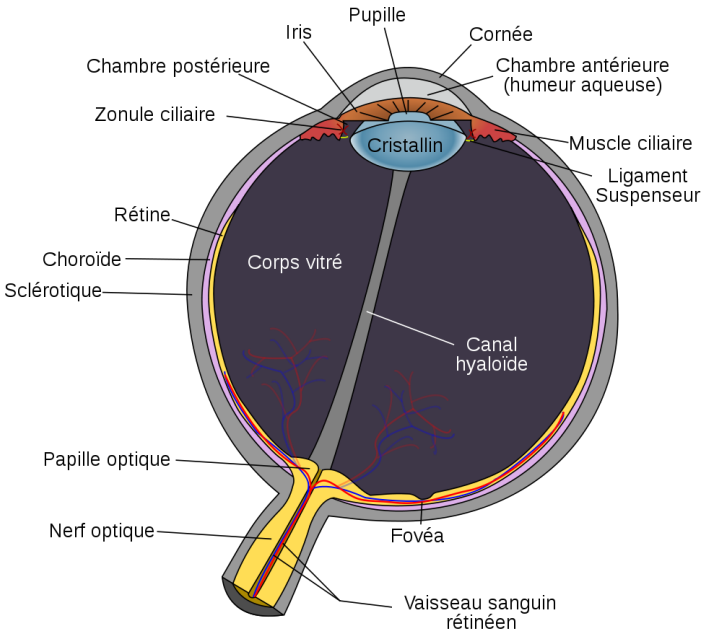


FIGURE 1.1 – Coupe de l'œil humain - Il possède trois membranes opaques (sclérotique, choroïde et rétine) et quatre milieux transparents (cornée, humeur aqueuse, cristallin et humeur vitrée).

L'iris

L'iris est une membrane pigmentée, perforée en son centre pour former la pupille. L'iris donne leur couleur aux yeux (en fonction de la pigmentation plus ou moins importante de la membrane).

Le cristallin

Le cristallin est une lentille bombée des deux côtés (biconvexe) située derrière l'iris. Elle est flexible, ce qui lui permet de modifier sa forme pour obtenir une image nette sur la rétine (phénomène d'accommodation). Avec le temps, les muscles responsables de la déformation du cristallin ont de moins en moins de puissance. La conséquence de cette perte de puissance est une perte de capacité d'accommodation. C'est ce qu'on appelle la presbytie. L'acuité visuelle n'est plus améliorable à ce stade.

L'humeur vitrée

L'humeur vitrée est le milieu transparent situé entre le cristallin et la rétine. Elle est principalement composée d'une substance gélatineuse contenant majoritairement de l'eau. En plus de laisser passer les rayons lumineux, l'humeur vitrée maintient le cristallin et la rétine en place en appliquant une pression sur les différentes membranes de l'œil. Tout

comme l'humeur aqueuse, elle se renouvelle de façon continue, ce qui permet l'évacuation des déchets du cristallin et de la rétine, en plus de les alimenter en glucose et en dioxygène.

La rétine

La rétine est une membrane épaisse constituée de plusieurs millions de cellules nerveuses photoréceptrices qui transmettent l'information visuelle vers le cerveau. Les cônes, qui permettent de percevoir la lumière du jour et les couleurs, et les bâtonnets qui nous donnent une vision des formes. Les cônes sont concentrés dans un petit cercle d'environ $2mm$ de diamètre, appelé la fovéa. La fovéa est la zone centrale de la macula. Cependant, il y a un endroit bien particulier de la rétine appelé la tache aveugle où il n'y aucun cône ni bâtonnet. En fait, c'est à cet endroit que commence le nerf optique. Les cônes et les bâtonnets sont liés à des cellules qui permettent à la rétine de transformer la lumière en impulsions électriques. La rétine se prolonge par le nerf optique qui transmet les impulsions jusqu'au cerveau.

Sur la FIGURE 1.2 [2], on peut voir que la concentration des bâtonnets et des cônes évolue selon l'endroit sur la rétine. Ils ne sont pas uniformément répartis. On voit qu'il y a un pic de concentration de cônes au niveau de la fovéa. On remarque aussi la présence d'une zone qui ne comporte ni cône, ni bâtonnet, c'est le point aveugle qui correspond à l'endroit où le nerf optique se dirige vers le cerveau.

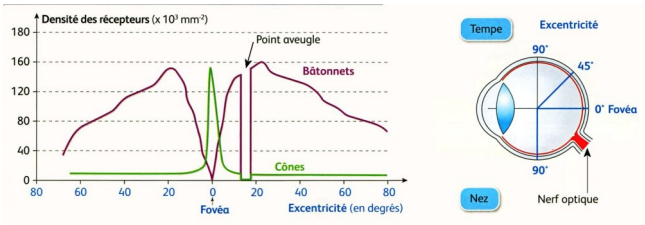


FIGURE 1.2 – Densité des photorécepteurs en fonction de l'excentricité sur la rétine de l'oeil gauche. Les cônes sont très localisés au niveau de la fovéa. Les bâtonnets sont répartis le long de la rétine. Au niveau du point aveugle (nerf optique), il n'y a pas de photorécepteurs [Rodieck, 2003].

Un montage à l'envers !

Lorsqu'on regarde de plus près le système optique au niveau de l'œil, on se rend compte que le "montage" est étrange, et semble monté à l'envers. Normalement, lors de la conception d'un système d'acquisition d'information, on met au premier plan le module d'acquisition, puis un moyen de transfert d'information (nerf optique dans notre cas) et enfin le dispositif de calcul (cortex).

La figure FIGURE 1.3 [3] montre une coupe simplifiée de l'œil qui met en évidence ce

2. https://svt07.wordpress.com/1-a-chapitre-1-loeil-du-message-lumineux-au-message-nerveux/photorecepteur-repartition/
3. https://www.allaboutvision.com/resources/retina.htm

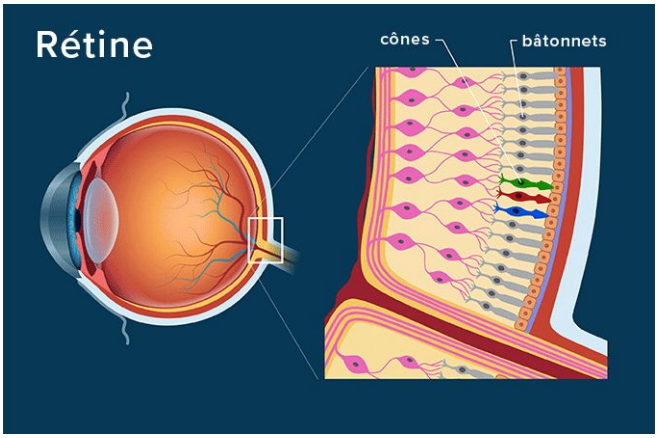


FIGURE 1.3 – Coupe simplifiée de l'œil. Les récepteurs (cônes et bâtonnets), sont placés au fond de la rétine, et communiquent avec les cellules ganglionnaires et le nerf optique en remontant vers l'amont.

phénomène. Les cônes et bâtonnets sont les cellules réceptrices du flux de lumière et elles sont placées derrière les cellules ganglionnaires qui communiquent avec le nerf optique.

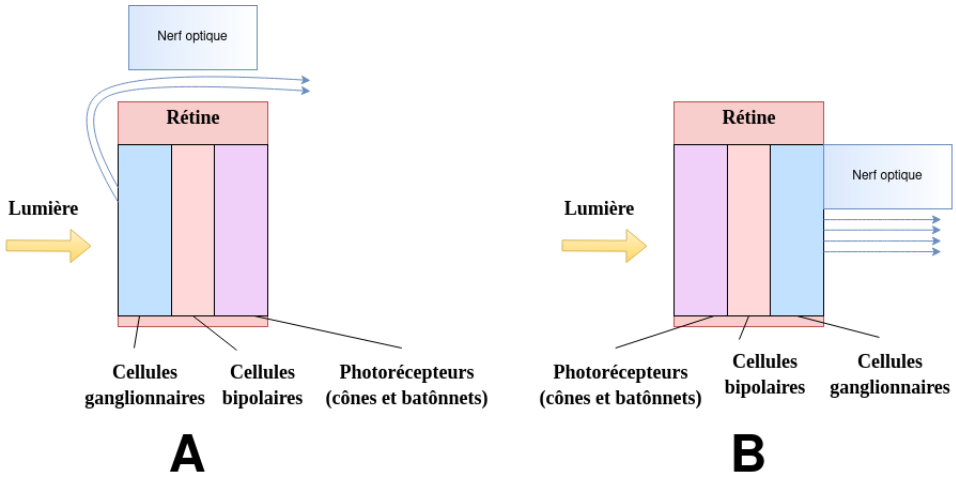


FIGURE 1.4 – Comparaison entre un schéma du système optique (œil et nerf optique) avec : - en A, le montage réel de l'œil (similaire à la FIGURE 1.3), - en B, le montage de l'œil s'il suivait un schéma proche de celui d'un appareil photo.

Si la rétine était montée comme un appareil photo alors un schéma représentant une coupe pourrait être celui montré sur la FIGURE 1.4. Si c'était le cas, il n'y aurait alors pas de tâche aveugle, étant donné que des photorécepteurs pourraient se trouver tout le long de la rétine.

1.1.2 Le nerf optique

Le nerf optique est le canal de communication entre les yeux (rétine) et le cerveau. C'est lui qui s'assure de la bonne communication des informations visuelles. Il est constitué de plus d'un million de cellules ganglionnaires. Le nerf optique issu de chaque œil se sépare en deux au niveau du chiasma optique afin d'alimenter chaque hémisphère du cortex. La FIGURE 1.5 illustre l'organisation interne des nerfs optiques.

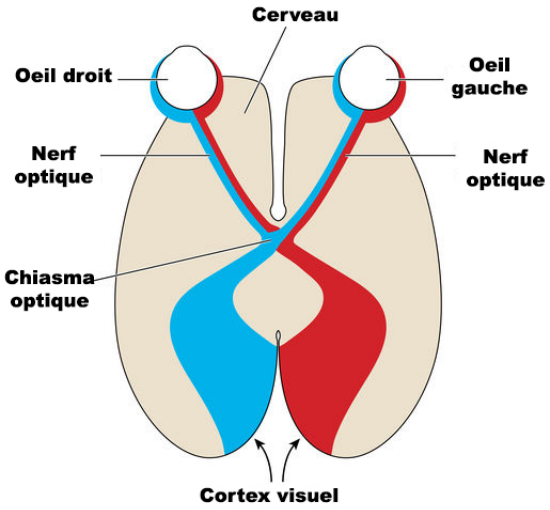


FIGURE 1.5 – Coupe transversale du cortex montrant la séparation du nerf optique au niveau du chiasma optique.

1.1.3 Le cortex visuel

Le cortex visuel est réparti à la fois sur les faces latérales et internes du lobe occipital et déborde légèrement sur le lobe pariétal et temporal. Son étude a permis de le découper en plusieurs sous-régions (*V1*, *V2*, *V3*, *V4*, *MT*, etc.) qui traitent chacune ou ensemble les informations qui arrivent depuis la rétine via le nerf optique (formes, couleurs, mouvements, etc.).

Carte rétinotopique

La carte rétinotopique, ou cartographie rétinotopique, est la façon de décrire l'adjacence de deux points voisins sur la rétine à l'adjacence de deux points sur le cortex. Dans notre cas, cette fonction de transition qui permet de passer de la rétine au cortex est très importante, car nous avons besoin de savoir comment une matrice d'électrodes sur la rétine va agir sur la représentation phosphénique perçue par le patient implanté. Les FIGURES 1.6 et 1.7 [4] illustrent le fonctionnement de la cartographie rétinotopique.

4. https://www.em-consulte.com/em/SFO/H2017/index.html

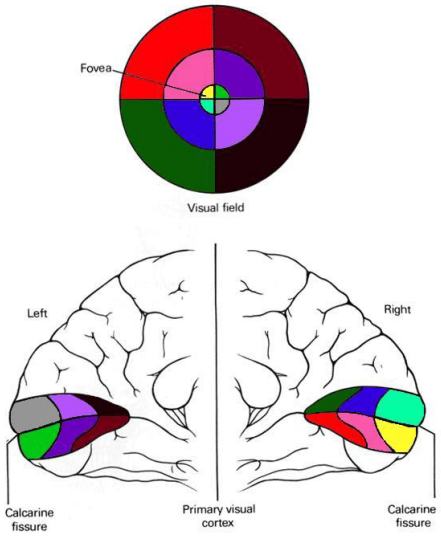


FIGURE 1.6 – Transformation du champ de vision (en haut de l'image), en zones stimulées dans le cortex visuel. Les couleurs correspondent aux zones stimulées. Extrait de [Calabrèse, 2011]

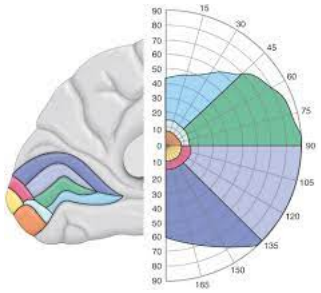


FIGURE 1.7 – Transformation du champ de vision (en haut de l'image), en zones stimulées dans le cortex visuel. Les couleurs correspondent aux zones stimulées. On montre ici le champ visuel droit tronqué ainsi que les différentes zones de vision colorées et retranscrites dans le cortex visuel gauche.

On remarque qu'en proportion, la fovéa occupe la plus grande zone stimulée du cortex.

Zones et traitements

Chaque zone du cortex visuel est liée au traitement d'un type d'information. L'aire *V1* est située autour de la scissure calcarine et est responsable de la détection des contrastes. Les aires *V2* et *V3* forment deux rubans corticaux, entourant la zone *V1*. Dans la zone *V2*, on distingue trois types de bandes : bandes fines, bandes épaisses et intermédiaires. Les bandes épaisses sont responsables de l'interprétation de l'orientation et de la direction du mouvement. Les bandes fines sont plutôt sensibles à la couleur. L'aire *V3* permet de

traiter les informations relatives aux mouvements et aux formes. Les FIGURES 1.8 [5] et 1.9 [6] montrent l'emplacement des différentes zones dans le cortex.

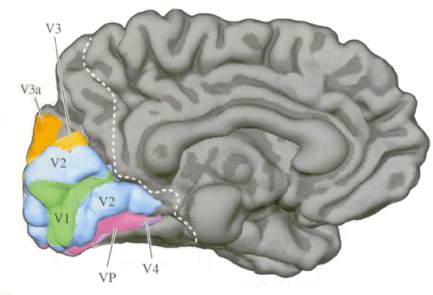


FIGURE 1.8 – Coupe sagittale de la face interne cortex. Les zones *V1*, *V2*, *V3*, *V4* et *V5* sont représentées dans le lobe occipital.

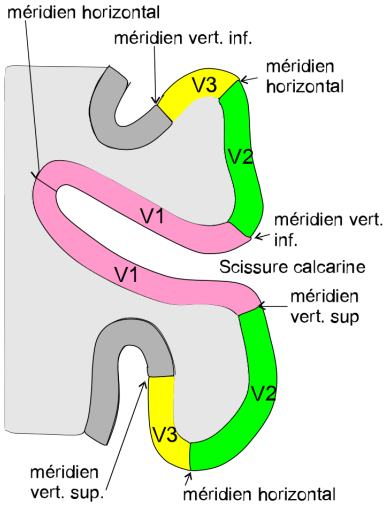


FIGURE 1.9 – Focus sur les zones *V1*, *V2* et *V3* du cortex. La zone *V1* se trouve dans la scissure calcarine et est un site privilégié pour la pose d'implants.

1.2 Déficience visuelle et caractérisation

La déficience visuelle correspond à une insuffisance ou une absence d'image perçue par le système visuel. Elle correspond à une atteinte de l'œil ou des voies visuelles jusqu'au système cérébral. Ces atteintes peuvent être congénitales ou acquises : accidents ou maladies, telles que le diabète, la DMLA ou le glaucome. La déficience peut porter sur l'acuité

5. http://acces.ens-lyon.fr/acces/thematiques/neurosciences/actualisation-des-connaissances/perception-sensorielle-1/vision/comprendre/vision_scientifique/organisation_cerebrale
6. https://fr.wikipedia.org/wiki/Cortex_visuel

visuelle (pourcentage restant par rapport à la vision normale) et sur le champ visuel, d'un œil ou des deux yeux.

1.2.1 Classification

L'OMS décrit six stades de déficience visuelle (classification internationale des handicaps), allant de la cécité totale à la vision presque normale (SNOF [7]) :

— la cécité absolue correspond à l'absence de perception de la lumière (catégorie 5) ;
— la déficience presque totale, ou cécité sévère, est définie par une acuité visuelle inférieure à 1/50 (ou comptage des doigts à un mètre ou moins) avec perception conservée de la lumière, ou un champ visuel inférieur à 5 degrés (catégorie 4) ;
— la déficience profonde correspond à une acuité visuelle inférieure à 1/20 (comptage des doigts à moins de 3 mètres) et supérieure à 1/50 ou un champ visuel inférieur à 10 degrés, mais supérieur à 5 degrés (catégorie 3) ;
— la déficience sévère correspond à une acuité visuelle inférieure à 1/10 et supérieure ou égale à 1/20 (catégorie 2) ;
— la déficience moyenne correspond à une acuité visuelle inférieure à 3/10 et supérieure ou égale à 1/10 avec un champ visuel d'au moins 20 degrés (catégorie 1) ;
— la déficience visuelle légère ou inexistante correspond à une acuité visuelle supérieure ou égale à 3/10 (catégorie 0).

Pour évaluer l'acuité visuelle, les ophtalmologues utilisent des grilles d'acuité représentant des lettres de différentes tailles. L'objectif pour le patient testé est de lire les lettres, de la police la plus grande à la plus petite, à une distance fixe de la grille. La dernière ligne lue donne le score d'acuité visuelle. La FIGURE 1.10 montre une grille utilisée par les opticiens pour évaluer l'acuité visuelle.

1.2.2 Dégénérescence maculaire liée à l'âge (DMLA)

La Dégénérescence Maculaire Liée à l'Âge [Massé and Buhannic, 2017] est une maladie dégénérative de la rétine d'évolution chronique qui touche les personnes de plus de 50 ans. Elle atteint la région maculaire, c'est-à-dire la zone centrale de la rétine, entraînant une perte progressive de la vision centrale, fine des détails (gêne pour la lecture de près, la conduite automobile...). Elle laisse habituellement intacte la vision périphérique ou latérale. Les patients atteints de DMLA ne deviennent donc jamais tout à fait aveugles puisqu'ils gardent leur vision périphérique. La FIGURE 1.11 illustre ce phénomène.

7. SNOF : Syndicat National des ophtalmologistes de France - https://www.snof.org/public/conseiller/malvoyance-et-handicaps-visuels

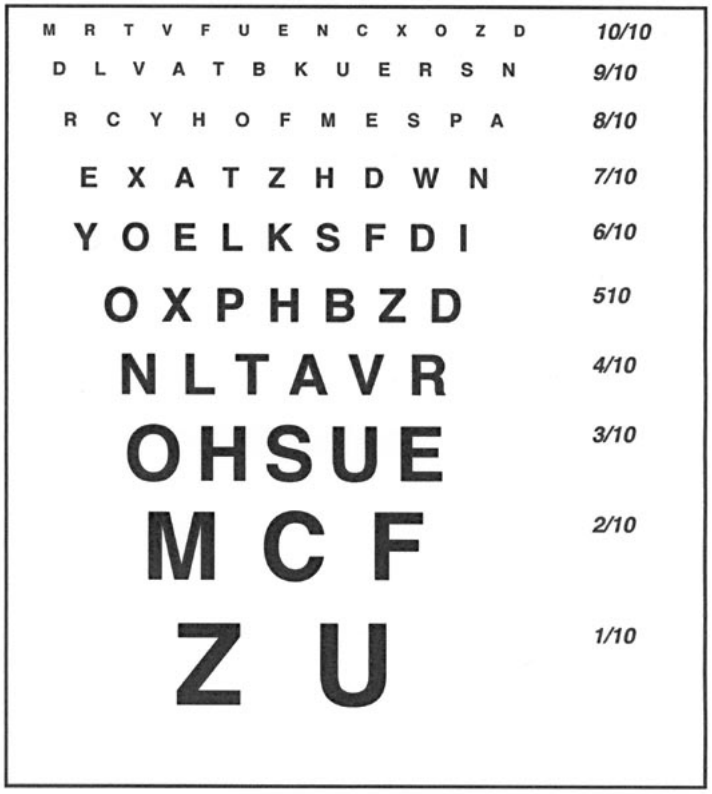


FIGURE 1.10 – Grille d'évaluation de l'acuité visuelle. La police des lettres est décroissante de bas en haut. Le sujet commence par lire les lettres les plus grosses. Une fois une ligne finie, il passe à celle du dessus. La dernière ligne lue sans difficulté fournit alors l'acuité visuelle.

1.2.3 Rétinite pigmentaire (RP)

La rétinite pigmentaire (RP) est également une maladie génétique dégénérative de l'œil qui se caractérise par une perte progressive et graduelle de la vision évoluant généralement vers la cécité. Elle touche environ 1 personne sur 3 500. C'est une maladie génétique due à l'altération de gènes impliqués dans le fonctionnement et la régulation des cellules de la rétine, les photorécepteurs, indispensables pour la vision.

La RP débute généralement par des problèmes de vue lorsque l'intensité de la lumière diminue. Des difficultés d'adaptation à l'obscurité sont fréquentes, par exemple lors du passage d'une pièce très éclairée vers une pièce sombre. Puis, progressivement, le champ visuel se rétrécit avec l'impossibilité de voir les choses en haut, en bas ou sur les côtés, donnant une impression de "vision en tunnel" : cela correspond à une réduction de la vision périphérique. Cette atteinte est bilatérale, les deux yeux sont affectés. Dans certains cas, d'autres problèmes oculaires peuvent s'ajouter à la RP. Par exemple, un œdème maculaire peut apparaître, et entraîner une baisse de la vision centrale avant le rétrécissement majeur du champ visuel. La FIGURE 1.11 illustre également ce phénomène.



FIGURE 1.11 – De gauche à droite, une image sans altération, la même image vue au travers d'un sujet atteint de DMLA et enfin la même image vue au travers d'un sujet atteint de RP.

1.3 Neuroprothèses visuelles

Une neuroprothèse visuelle est un dispositif permettant à des non-voyants de percevoir leur environnement grâce à une micro-stimulation électrique du système visuel. Dans la suite de cette section, toutes les prothèses présentées seront basées sur une stimulation via des électrodes.

Les neuroprothèses visuelles sont généralement constituées de trois éléments. Une caméra vidéo dont la fonction est de convertir la lumière en flux vidéo. Ce flux vidéo est ensuite traité par un micro-ordinateur portable. Un signal post-calcul (décrivant la stimulation des électrodes) est ensuite transmis à la matrice d'électrodes, appelée MEA (*multi-electrode array*), qui est en contact avec les tissus nerveux.

Il existe différents sites de fixation pour la matrice d'électrodes le long du système optique. Une matrice placée en amont du système (rétine) permet d'utiliser une grande quantité du système optique fonctionnel du patient, mais peut entraîner la non-transmission ou l'altération des signaux. Une matrice située dans une portion tardive ne bénéficie pas de tout le système optique fonctionnel, mais permet de réduire la dégradation du signal.

Il existe quatre sites d'implantation le long du système optique, détaillés dans la suite de cette section. On trouve les implants rétiniens (1.3.1), corticaux (1.3.2), ceux placés sur le corps géniculé latéral (1.3.3) et ceux placés le long du nerf optique (1.3.4). Un tableau récapitulatif des différents implants est ensuite proposé (1.3.5).

1.3.1 Neuroprothèses rétiniennes

Depuis presque 30 ans, les efforts mis en place dans le développement des prothèses rétiniennes ont abouti à la commercialisation de plusieurs dispositifs qui ont montré leur intérêt dans des tâches simples de la vie quotidienne [Bloch et al., 2019, Velázquez, 2017, Conradie et al., 2014]. Il existe 3 types de neuroprothèses rétiniennes : les implants épirétiniens qui sont placés contre la rétine à l'intérieur de l'œil et sont donc en contact avec les cellules ganglionnaires qu'ils stimulent, les implants sous-rétiniens qui sont placés directement sur les cellules photoréceptrices endommagées et les implants supra-choroïdiens

qui sont fixés sous la sclérotique, contre la choroïde. La FIGURE 1.12 montre les différents sites d'implantation.

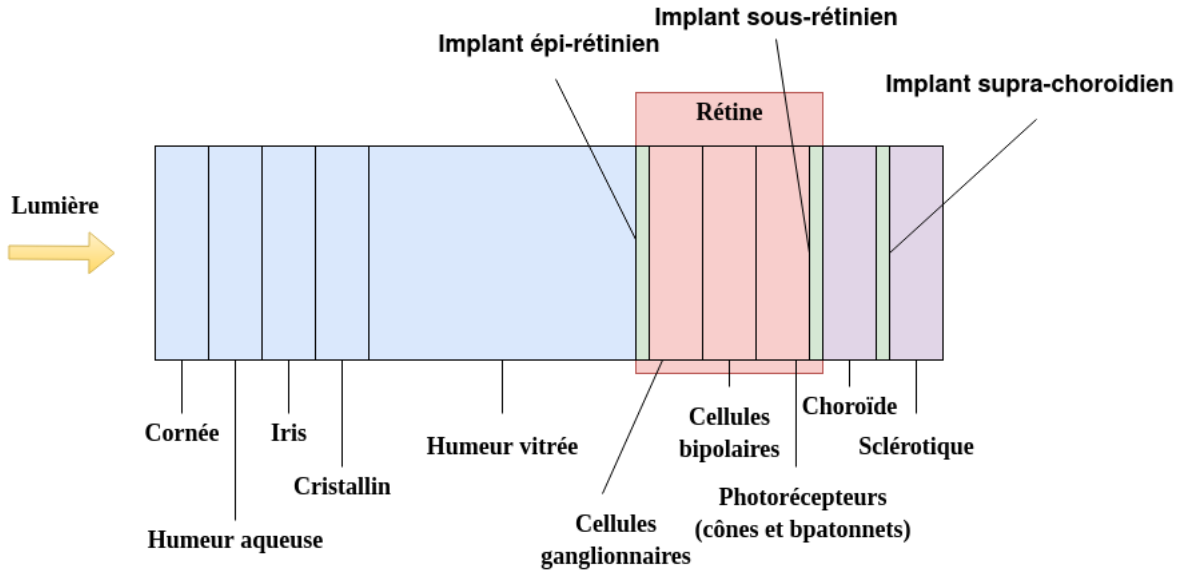


FIGURE 1.12 – Schéma simplifié de l'œil avec les différents sites d'implantations de neuroprothèses rétiniennes (en vert). La zone rouge correspond aux différentes couches de la rétine. Les implants peuvent être épirétiniens, sous-rétiniens ou supra-choroïdiens.

Neuroprothèses épirétiniennes

Comme mentionné ci-dessus, les implants épirétiniens sont insérés à la surface de la rétine, au contact des cellules ganglionnaires (dont les axones forment le nerf optique). Le réseau d'électrodes est fixé à la surface de la rétine à l'aide d'une punaise. Ce type de prothèse présente deux avantages : le site d'implantation est l'un des plus faciles à atteindre pour les chirurgiens compte tenu de sa localisation, et il permet également de mieux diffuser la chaleur générée par la matrice d'électrodes.

Elles présentent cependant des inconvénients : la stimulation des cellules ganglionnaires de la rétine contourne le chemin classique de l'information visuelle qui doit normalement partir des récepteurs (cônes et bâtonnets) vers les cellules ganglionnaires. Il est donc difficile de reconstruire fidèlement la cartographie rétinotopique. De plus, étant donné la proximité de la matrice avec les axones du nerf optique, des stimulations involontaires peuvent entraîner l'apparition de phosphènes indésirables. Des études expérimentales ont démontré que des phosphènes peuvent être induits chez 95% des sujets par une stimulation épi-rétinienne [Keserü et al., 2012].

Argus II

Le dispositif Argus II (Second Sight Medical Products Inc., Sylmar, CA, USA) est un implant épirétinien composé de deux parties, elles-mêmes composées de plusieurs éléments. Le premier composant (externe) comporte des lunettes équipées d'une caméra, d'un dispositif de calcul et d'une bobine de communication émettrice. Le second composant (interne) est composé d'une bobine de communication réceptrice, ainsi que d'un boîtier de décodage du signal et enfin de la matrice d'électrodes. La FIGURE 1.13 illustre les deux composants.

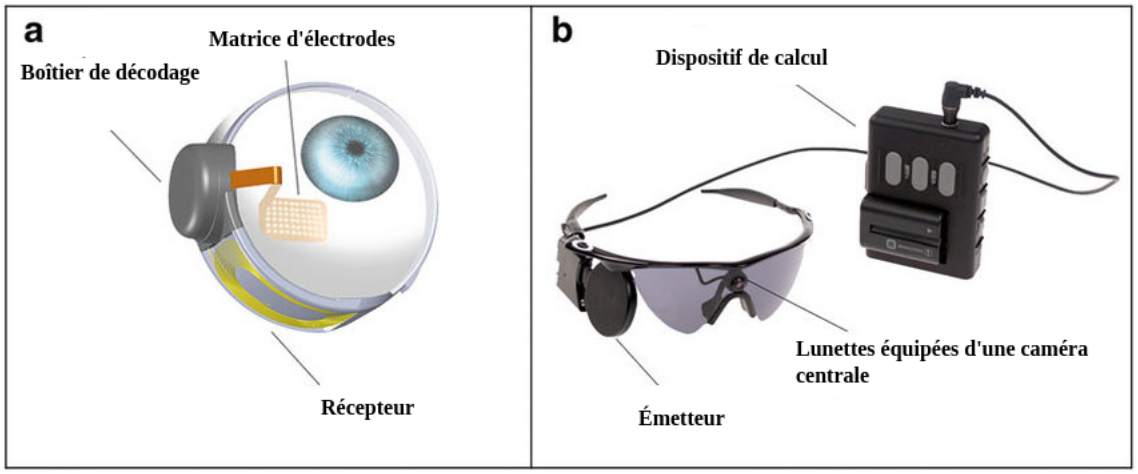


FIGURE 1.13 – Neuroprothèse Argus II - La figure montre les deux composants de la neuroprothèse. Sur la partie gauche (a), on retrouve les éléments internes de la neuroprothèse (matrice d'électrodes, bobine réceptrice et boîtier de décodage). Sur la partie (b), on retrouve les éléments externes de la neuroprothèse (dispositif de calcul, lunettes équipées d'une caméra, émetteur). Image adaptée de [Dagnelie et al., 2017]

Le flux d'images acquis par la caméra est ensuite envoyé à une première boîte de calcul qui construit le signal de stimulation de la matrice d'électrode. Une fois calculé, ce signal est converti et envoyé au boîtier de décodage à travers les deux bobines communicantes (émetteur/récepteur). Une fois le signal décodé, la stimulation micro-électrique est appliquée à la matrice d'électrodes générant l'apparition de phosphènes. La matrice de stimulation est composée de 60 micro-électrodes (organisées sur une grille 6x10), chacune de 200 μm de diamètre, espacées de 525 μm, couvrant un champ de vision de 20 degrés. Le système interne a démontré une durée de vie de plus de 10 ans lors de tests de vieillissement accéléré.

Au cours des différentes phases des essais cliniques réalisés sur les sujets (environ 30 sujets), il a été montré que ces derniers obtenaient de meilleures performances en matière d'acuité visuelle, de localisation de formes géométriques et de direction du mouvement lorsque le dispositif était activé que lorsqu'il était désactivé. Afin d'évaluer les performances des sujets, montrées sur la FIGURE 1.14, cinq tâches ont été mises en place (les trois premières sans navigation et les deux suivantes avec navigation) :

— une tâche de localisation d'un carré lumineux blanc sur un fond noir évaluée en comparant la distance entre le point touché par le sujet et la position réelle du carré (FIGURE 1.14 - a) ;
— une tâche dans laquelle le sujet doit déterminer la direction du mouvement d'une barre blanche qui se déplace sur un fond noir, évaluée en mesurant l'écart angulaire entre la direction donnée par le sujet et la direction réelle de la barre (FIGURE 1.14 - b) ;
— une tâche d'identification de motifs, alternant des bandes noires et blanches, où le patient doit déterminer le nombre de bandes noires (FIGURE 1.14 - c) ;
— une tâche de suivi de ligne blanche au sol évaluée en mesurant le temps où les sujets restent sur cette dernière (FIGURE 1.14 - d) ;
— une tâche de détection de porte où les sujets doivent se déplacer depuis un point initial vers une porte blanche sur un mur noir et étant évalués sur le nombre de réussite (FIGURE 1.14 - e).

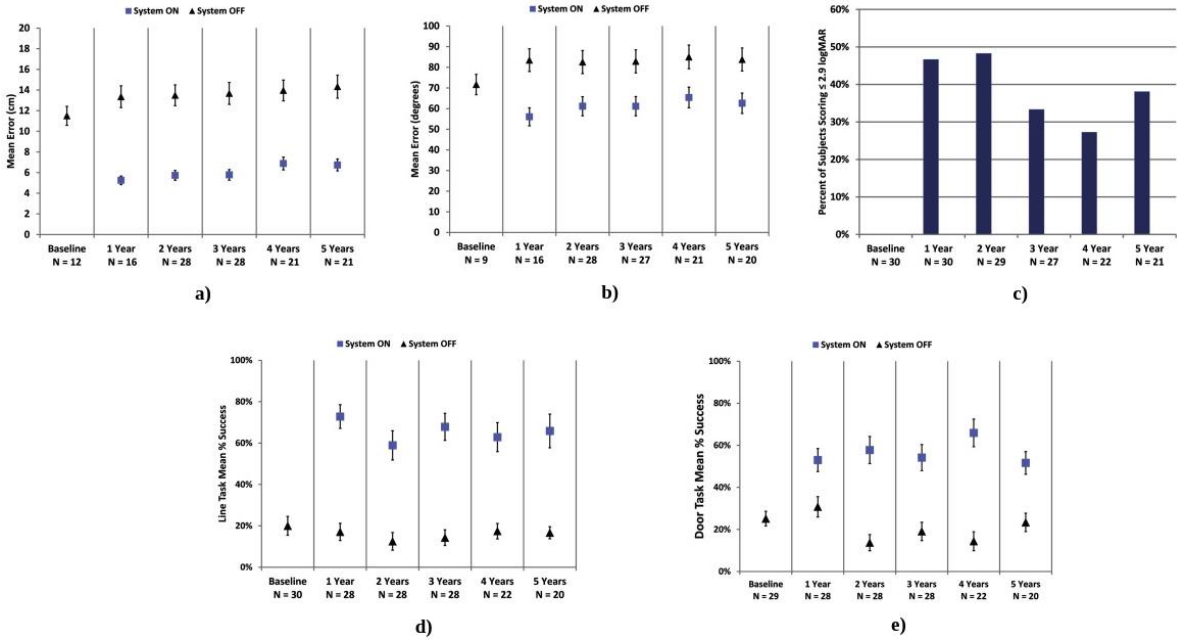


FIGURE 1.14 – Résultats des cinq tâches conduites par da Cruz et al. en 2016. Les figures montrent l'évolution des performances des 30 sujets sur cinq ans avant implantation (colonne "*baseline*"), puis sur cinq ans après implantation (colonnes suivantes). Les rectangles bleus correspondent aux performances lorsque le système Argus II est allumé, et les triangles noirs lorsqu'il est éteint. a) Erreur moyenne de localisation d'un carré blanc sur fond noir ; b) Erreur angulaire moyenne de direction d'une barre mobile ; c) Pourcentage des sujets qui obtiennent un score d'acuité visuelle supérieur à un seuil donné ; d) Pourcentage du temps passé par les sujets le long d'une ligne blanche au sol ; e) Pourcentage de réussite dans l'atteinte d'une porte blanche sur un mur noir.

Toutes ces évaluations ont été réalisées avec l'Argus II allumé et éteint. Les résultats montrent par exemple que dans le cas de la tâche de localisation du carré, les sujets réalisent des performances en moyenne deux fois meilleures avec le système allumé. De plus, les performances des sujets se sont maintenues au cours des 5 années suivant l'implantation [da Cruz et al., 2016].

Des tâches de lecture de lettres [Dagnelie et al., 2017], de préhension [Kotecha et al., 2014], et de navigation [Dagnelie et al., 2017] ont toutes montré une différence significative avec le dispositif activé chez les patients avec l'Argus II. Actuellement, un programme d'évaluation médicale a estimé que davantage de données sont nécessaires pour vérifier si les avantages pour le patient en termes de qualité de vie et d'activités de la vie quotidienne sont suffisamment élevés.

L'Argus II est le premier dispositif à recevoir le marquage de protection européen (CE) en 2011, puis l'approbation de la *Food and Drug Administration* (FDA) en 2013. Ces deux validations impliquent une conformité du produit aux règles de santé établies à l'échelle européenne et américaine. Il s'agit de la prothèse rétinienne la plus utilisée au monde, avec plus de 250 patients à avoir été implantés à ce jour.

IRIS II

L'IRIS II est une amélioration du modèle précédent appelé IMI, suite au rachat de ce dernier par la société Pixium Vision S.A. en 2007. IRIS est l'acronyme de "*Intelligent Retinal Implant System*". Les tests réalisés sur l'IMI montrent des résultats satisfaisants en termes de sécurité et de longévité, avec des phosphènes et des motifs rapportés par les patients lors de la stimulation [Keserü et al., 2012].

L'appareil est composé d'un mécanisme d'acquisition d'images et d'un mécanisme de transmission d'informations par infrarouge. Il comprend également une boîte de calcul qui permet de construire les commandes de stimulation à envoyer à la matrice d'électrodes via un dispositif infrarouge, voir Figure 1.15 [8]. La matrice contient 150 micro-électrodes. Le mécanisme utilise un capteur spécial (capteur neuromorphique) pour faire une acquisition plus rapide des signaux d'entrée (perception de l'environnement).

Les capteurs neuromorphiques se basent sur la variation d'intensité des pixels captés et ne propagent dans le dispositif de calcul que cette différence. La raison pour laquelle il s'agit d'une bonne technique, réside dans la limitation de la quantité d'informations envoyées et donc dans une réduction de la résolution temporelle de la rétine, tout en réduisant le volume d'informations visuelles redondantes lors du calcul de la stimulation. Le taux de transfert des données est donc augmenté. La plus grande force de cet implant, est qu'il inclut un module d'apprentissage rétinien, qui permet un étalonnage individualisé, afin d'assigner des zones comme excitatrices ou inhibitrices de la rétine [Eckmiller et al., 2005].

8. https://www.pixium-vision.com/fr

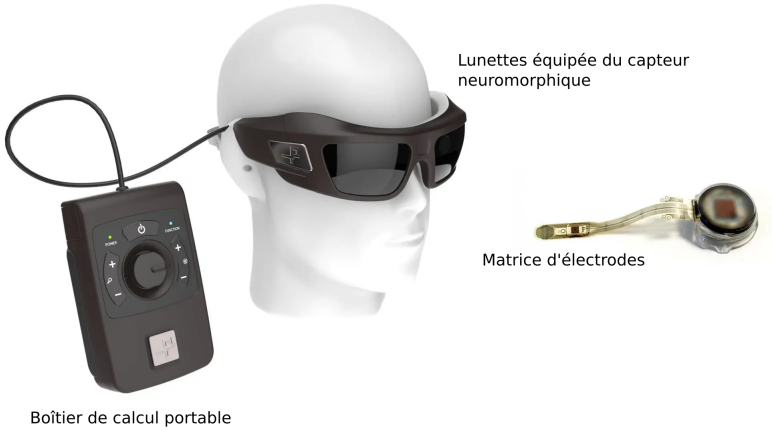


FIGURE 1.15 – Neuroprothèse IRIS II - La figure montre les différents composants de la neuroprothèse (boîtier de calcul, lunettes et matrice d'électrodes).

L'IRIS II a reçu l'approbation CE en 2016 et les premiers résultats sont prometteurs [Muqit et al., 2019]. Des tâches similaires à celles réalisées dans le cas de l'Argus II ont été conduites. On retrouve notamment une tâche de localisation d'un carré blanc sur fond noir, une tâche de détection de direction du mouvement et une tâche évaluant l'acuité visuelle. À cela s'ajoute une tâche de reconnaissance d'image. La FIGURE 1.16 montre les résultats obtenus. La commercialisation n'est pas encore à l'ordre du jour, car Pixium Vision a reporté l'essai dans l'attente d'un perfectionnement du dispositif, la durée de vie de l'implant étant inférieure aux estimations.

EPI-RET 3

L'EPI-RET3 comprend, comme les deux autres prothèses précédentes, deux modules. Le module interne est, pour cet implant, contrairement à l'Argus II et l'IRIS II entièrement intra-oculaire. Il est composé d'une bobine réceptrice, d'une puce électrique et de la matrice d'électrodes. Il n'y a pas de câble sortant du globe oculaire. Ce mécanisme limite les risques d'infection ou d'érosion de l'œil. Le module externe comprend une caméra et un processeur, qui transmet sans fil la carte phosphénique calculée sous forme d'impulsions qui stimulent les composants internes, voir FIGURE 1.17.

L'EPI-RET3 utilise un mécanisme d'impulsions à haute fréquence pour transmettre les informations à la puce intraoculaire et limiter les artefacts de stimulation. Les expérimentations animales ont montré la stimulation des cellules ganglionnaires autour des électrodes stimulées, démontrant une bonne efficacité du mécanisme. Le système est similaire à celui prévu pour l'humain.

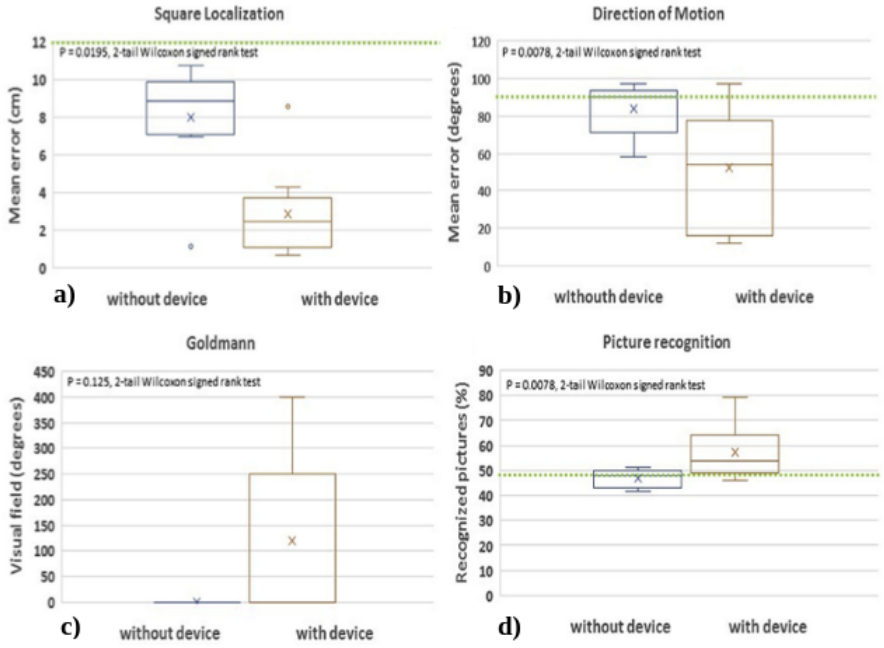


FIGURE 1.16 – Résultats des quatre tâches conduites par Muqit et al. en 2019 sur l'IRIS II. Les tests sont réalisés six mois après l'implantation. Les diagrammes bleus correspondent aux cas où le dispositif est éteint, les diagrammes oranges aux cas où il est allumé. Les lignes vertes pointillées correspondent au seuil de chance. a) Erreur moyenne de localisation d'un carré blanc sur fond noir ; b) Erreur angulaire moyenne de direction d'une barre mobile ; c) Champ visuel mesuré sans ou avec le dispositif allumé ; d) Pourcentage d'images reconnues sans ou avec le dispositif allumé.

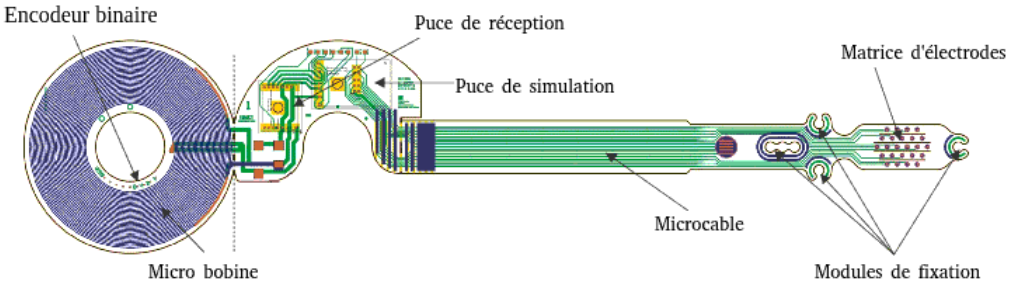


FIGURE 1.17 – Neuroprothèse EPI-RET 3 - La figure montre les différents composants de la neuroprothèse. Source : Équipe EPI-RET3, RWTH Aachen, Allemagne.

Neuroprothèses sous-rétiniennes

Les prothèses sous-rétiniennes se caractérisent par leur placement juste sous la rétine, au niveau des photorécepteurs endommagés. Ce procédé est intéressant, car il permet d'exploiter pleinement la capacité de traitement du signal des interneurones rétiniens, produisant une forme de vision plus physiologique avec moins de traitement d'image. De

plus, le dispositif étant plus proche des cellules d'acquisition, il nécessite des intensités de stimulation plus faibles.

Ce type d'implant ne peut être utilisé que si le réseau de cellules optiques est fonctionnel, ce qui est très peu probable en cas de déficience visuelle [Jones and Marc, 2005]. En termes de contraintes d'implantation, il existe deux difficultés principales, la première étant que l'approche chirurgicale est moins familière aux chirurgiens pratiquant la chirurgie rétinienne de routine. La seconde difficulté est due au fait que la chirurgie nécessite la séparation de l'épithélium pigmentaire rétinien et de la rétine pour insérer l'implant.

BOSTON RETINAL IMPLANT (BRIP)

Le "*Boston Retinal Implant Project*" (BRIP) a été l'un des premiers projets à conduire des essais cliniques sur des humains. Il a été démontré que des percepts reproductibles pouvaient être induits par la stimulation d'une électrode unique chez des patients atteints de rétinite pigmentaire au stade terminal et chez un patient ayant une vision normale, mais un cancer de l'œil [Grumet et al., 2000]. Le dispositif BRIP est très proche en termes de conception de l'implant Argus II, mais il est implanté dans l'espace sous-rétinien, afin d'éviter la nécessité de fixer le dispositif et de minimiser les complications post-opératoires.

Les équipes n'ont cependant pas encore implanté de manière définitive un sujet non-voyant et se concentrent sur le développement d'un implant 16x16 (256 électrodes), seuil qui, selon eux, est nécessaire pour rendre une vision fonctionnelle [Kelly et al., 2011, Rizzo III, 2011]. Les FIGURES 1.18 et 1.19 illustrent l'implant.

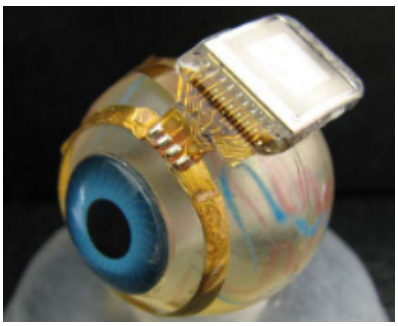


FIGURE 1.18 – Neuroprothèse BRIP - L'implant entoure l'œil et la matrice d'électrodes se trouve dans l'espace sous-rétinien.

ARTIFICIAL SILICON RETINA (ASR)

Ce dispositif fait partie de la catégorie de prothèses utilisant la lumière ambiante pour stimuler la rétine. La matrice de rétine artificielle en silicium, d'une largeur de 2 mm et d'une épaisseur de 25 μm, est constituée de 5 000 micro-photodiodes de 20 μm de diamètre associées à des micro-électrodes de 9 μm de diamètre à pointe d'iridium et séparées par 5 μm. Dans une étude réalisée sur six patients, il a été démontré que des phosphènes

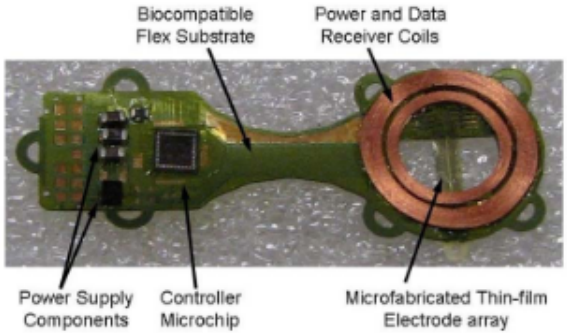


FIGURE 1.19 – Neuroprothèse BRIP - L'implant est flexible et s'enroule autour de l'œil. Il est composé de plusieurs bobines et de la matrice d'électrodes.

pouvaient être perçus dans la région du champ de vision correspondant au dispositif chez quatre patients sur six [Chow et al., 2004]. Quatre autres patients ont été implantés, ce qui donne un total de 10 patients implantés avec la puce ASR, dont 6 ont bénéficié d'un suivi à long terme [Chow et al., 2010]. Bien qu'une amélioration temporaire de la fonction visuelle généralisée ait été démontrée, par rapport à un œil témoin, elle a été attribuée à l'effet des facteurs de croissance neuroprotecteurs plutôt qu'à la prothèse rétinienne en soi. Il a été conclu qu'un dispositif reposant uniquement sur la lumière ambiante est incapable de générer un courant suffisant pour stimuler directement un nombre significatif de neurones. L'entreprise a par la suite fermé ses portes.

ALPHA IMS

L'Alpha IMS est le premier et le seul implant sous-rétinien à avoir obtenu le marquage CE, qui a été accordé en 2013. De manière similaire à l'ASR, il intègre un réseau photovoltaïque, appelé réseau multiphotodiode, qui comprend une micro-puce de 3 mm^2 contenant 1 500 unités "photodiode-amplificateur-électrode" indépendantes, chacune d'entre elles convertissant la luminance ambiante en un signal électrique. L'implant couvre une surface de 9 mm^2, ce qui correspond à un champ visuel compris entre 15° et 25°. Il a une épaisseur de 70 μm [Zrenner et al., 2009]. Cependant, il diffère de l'ASR dans le sens où il s'agit d'un dispositif dit "actif", utilisant une source d'énergie supplémentaire pour amplifier le signal. Une bobine externe amovible se fixe magnétiquement à la bobine sous-dermique, permettant l'induction d'énergie électromagnétique et le contrôle de la sensibilité au contraste et de la luminosité [Zrenner et al., 2011, Zrenner et al., 2017]. Les FIGURES 1.20 et 1.21 illustrent respectivement l'ensemble de la neuroprothèse sur le système visuel humain et un focus sur la matrice d'électrodes.

Une première étude en 2013 a montré que sur 9 patients, 8 parvenaient à détecter la lumière, 7 arrivaient à la localiser et 5 étaient capables de détecter des mouvements [Stingl et al., 2013]. Un deuxième essai clinique de l'Alpha IMS réalisé en 2014 visait à évaluer l'amélioration de la vie quotidienne et de la mobilité, ainsi que l'acuité visuelle et

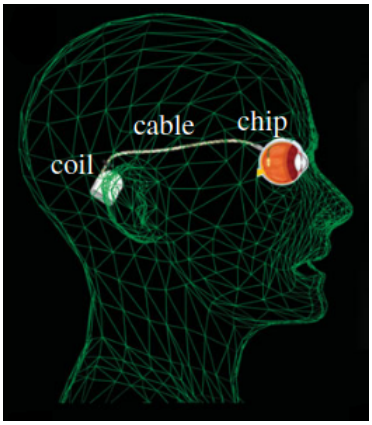


FIGURE 1.20 – Neuroprothèse ALPHA IMS - La bobine sous-dermique située derrière l'oreille fournit de l'énergie et envoie des signaux de commande à la matrice d'électrodes située dans l'œil via un câble sous-dermique. Extrait de [Stingl et al., 2013].

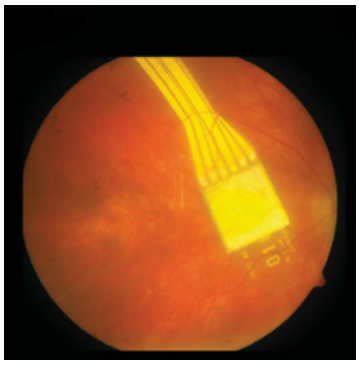


FIGURE 1.21 – Neuroprothèse ALPHA IMS - La matrice est placée chirurgicalement sous la fovéa et contient 1500 électrodes. Extrait de [Stingl et al., 2013].

la reconnaissance des objets. La meilleure acuité visuelle enregistrée était de 20/546. Les tests de reconnaissance d'objets ont révélé une amélioration significative avec l'appareil en place au cours des 3 premiers mois, mais sont tombés en dessous du seuil de signifiance à partir du sixième mois [Stingl et al., 2015]. Le profil de sécurité de l'appareil a été généralement considéré comme cliniquement acceptable [Kitiratschky et al., 2015].

SYSTEME PRIMA

Ce modèle d'implant est proposé par une société française du nom de Pixium Vision. C'est le modèle d'implant que nous avons choisi de simuler dans la plupart de nos expérimentations. Dans ce dispositif, une puce hexagonale de 1 mm de large, contenant 142 cellules de 30 μm d'épaisseur, est insérée par voie sous-rétinienne. Chaque pixel reçoit des informations visuelles sous forme de lumière pulsée dans le proche infrarouge, directement à partir d'une paire de lunettes spécialement conçues. Le signal passe ensuite d'une

électrode stimulante à une électrode de réception connectée à plusieurs photodiodes. Ces photodiodes génèrent un courant électrique qui polarise le tissu neuronal adjacent. Cette disposition modulaire est censée améliorer la résolution spatiale et être plus facilement adaptable à un champ visuel plus large. La FIGURE 1.22 illustre la composition et le fonctionnement de la neuroprothèse.

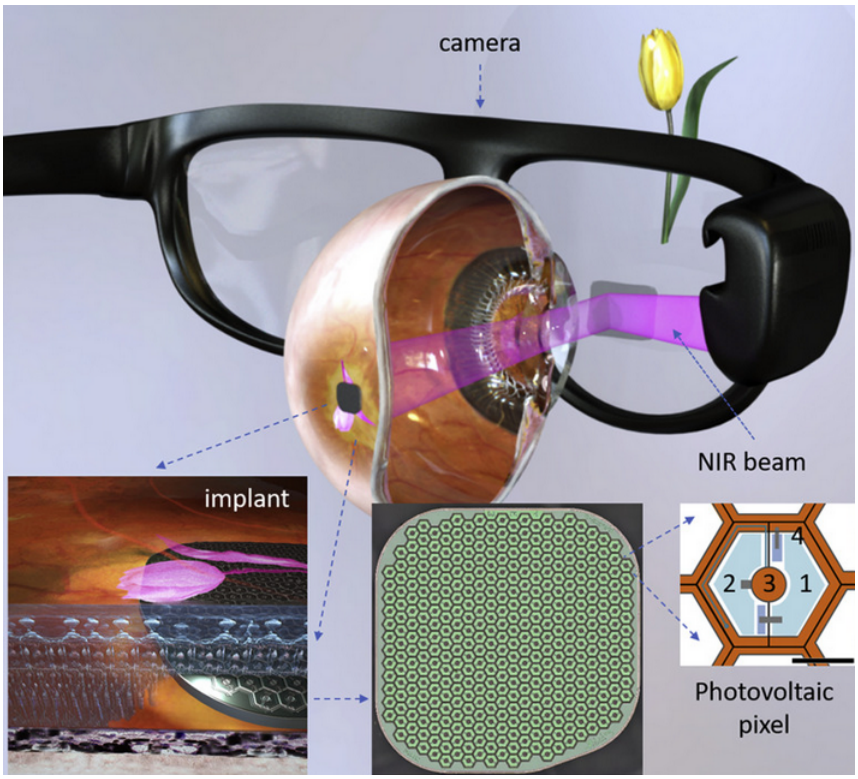


FIGURE 1.22 – Neuroprothèse PRIMA - Schéma du système montrant la composition et le fonctionnement du système PRIMA. Un réseau photovoltaïque sans fil sous-rétinien convertit la lumière pulsée en courant électrique pulsé dans chaque pixel pour stimuler les neurones internes adjacents de la rétine. Chaque pixel comprend 2 diodes (1 et 2), connectées en série entre les électrodes active (3) et de retour (4).

Les résultats précliniques de cette approche sur des modèles animaux sont encourageants [Mathieson et al., 2012, Mandel et al., 2013]. Cinq patients atteints de dégénérescence maculaire liée à l'âge (DMLA) sèche ont été implantés avec le dispositif PRIMA en France, et cinq autres aux États-Unis, dans le cadre d'une étude de faisabilité d'évaluation de la sécurité et des performances sur 36 mois. Ces données de suivi à long terme montrent que l'implant a été bien toléré et sûr pour les patients 36 mois après l'implantation. L'implant s'est aussi montré très stable, puisqu'aucune détérioration n'a été constatée au cours de sa durée de vie. L'étude de faisabilité a par ailleurs permis, après un processus de rééducation, d'améliorer considérablement l'acuité visuelle des patients [Palanker et al., 2020]. À la suite de ces résultats positifs, Pixium a démarré une étude européenne impliquant 38 patients,

sur trois années de suivi. Les premiers résultats de l'étude devraient être publiés en 2023, en vue d'une possible mise sur le marché début 2024.

Neuroprothèses suprachoroïdales

Le troisième type d'implant rétinien est le suprachoroïdien. Ces implants sont localisés entre la sclérotique et la choroïde. Sur ce site d'implantation, le système ne nécessite pas de chirurgie très invasive et est par conséquent plus facilement accessible pour une réparation ou un remplacement. Cependant, l'espace suprachoroïdien est très vascularisé et il existe un risque important d'hémorragie et un risque de fibrose après l'implantation. De plus, en raison de sa distance par rapport à la rétine, ce modèle semble nécessiter une plus grande puissance de stimulation pour susciter des perceptions visuelles. La localisation suprachoroïdienne risque également une plus grande diffusion du courant, réduisant ainsi la résolution spatiale.

BIONIC VISION AUSTRALIA

L'équipe de Bionic Vision Australia (BVA) a développé une série de prototypes d'implants suprachoroïdiens au cours des 10 dernières années. Le premier d'entre eux était un système à 24 électrodes, divisées en 20 électrodes de stimulation et 4 électrodes de retour. Ce dispositif n'a pas de propriétés photovoltaïques, s'appuyant sur une caméra montée sur la tête et un processeur d'images pour fournir les modèles de stimulation des électrodes. En 2012, trois sujets atteints de rétinite pigmentaire avancée ont été implantés pendant 2 ans dans le cadre d'une étude pilote. Pendant les tests, l'emplacement, la forme et la taille des phosphènes ont été cartographiés à l'aide d'un dispositif de suivi de mouvement monté sur le doigt et d'une caméra orientée vers l'œil pour surveiller le regard. Les phosphènes ont pu être déclenchés chez tous les patients [Ayton et al., 2014, Sinclair et al., 2016]. Des tests de localisation de la lumière ainsi que des tâches de la vie quotidienne ont été effectués à l'aide d'une caméra montée sur la tête. La localisation de la lumière a été supérieure chez tous les participants, tandis qu'un seul sujet a réussi le test d'acuité visuelle, avec une moyenne de 20/8397, nettement meilleure avec le dispositif en place qu'avec le dispositif éteint [Ayton et al., 2014]. Les principales limites de la stimulation suprachoroïdienne sont liées à la proximité du dispositif avec les neurones rétiniens. La FIGURE 1.23 illustre la composition et le fonctionnement de la neuroprothèse BVA.

Le dispositif de nouvelle génération que BVA développe devrait comporter à 44 électrodes et est déjà à l'étude [Abbott et al., 2018]. Ils développent également un dispositif à 99 électrodes, le Phoenix-99 [Suaning et al., 2014].

STIMULATION SUPRACHOROIDAL TRANSRETINAL

Le système de stimulation suprachoroïdienne-transrétinienne (STS) est actuellement en cours de développement (NIDEK). Le mode de fonctionnement est similaire à celui du

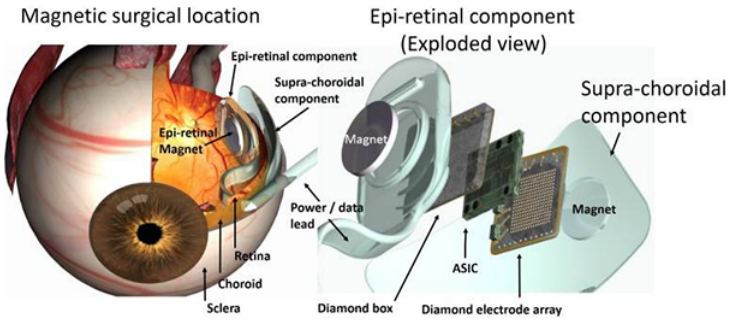


FIGURE 1.23 – Neuroprothèse BVA - Schéma du système montrant la composition et le fonctionnement de la neuroprothèse BVA. La matrice d'électrodes comporte 24 électrodes.

BVA. Une fois que la lumière est détectée par une caméra montée sur des lunettes, elle est traitée par un ordinateur situé dans la branche des lunettes, la bobine externe relaie un signal via la bobine secondaire au décodeur qui, à son tour, génère une impulsion pour stimuler les électrodes individuelles. Contrairement à d'autres systèmes, le STS actuel est constitué d'un réseau de 49 micro-électrodes "3D" (tridimensionnelles). Dans une étude pilote menée sur deux patients utilisant un prototype d'implant à neuf électrodes, il a été démontré que des phosphènes pouvaient être provoqués de manière reproductible dans la zone du champ visuel correspondant à l'implant, lors d'une stimulation directe.

Nous avons présenté les neuroprothèses rétiniennes. La sous-section suivante présente les différentes neuroprothèses corticales.

1.3.2 Neuroprothèses corticales

L'une des caractéristiques les plus importantes des prothèses corticales est leur localisation en aval du système optique. Cela offre un potentiel de réhabilitation pour les personnes aveugles et malvoyantes pour lesquelles une prothèse visuelle de type rétine, nerf optique ou CGL serait inefficace [Gabel, 2017].

GENNARIS

Monash Vision Group (MVG) est un partenariat de collaboration basé à Melbourne entre l'Université Monash et Alfred Health. L'objectif principal du MVG est de développer une prothèse de vision corticale, cliniquement viable, connue sous le nom de système de vision bionique Gennaris. Le Gennaris a été conçu pour contourner les dommages causés à l'œil et au nerf optique, afin de redonner une vision fonctionnelle aux personnes qui ont perdu la vue à la suite d'un traumatisme ou de maladies telles que le glaucome et les maladies rétiniennes acquises. Le système Gennaris est composé de 6 tuiles en céramique contenant chacune 43 électrodes, soit un total de 258 électrodes. Un processeur de poche

conçu sur mesure extrait les caractéristiques utiles de l'image capturée par la caméra et génère un modèle d'activation des électrodes afin de créer la carte de stimulation phosphénique [Lowery et al., 2015, Lowery et al., 2017]. Ils sont les premiers à introduire le concept de "*transformative reality*" (TR) qui, au lieu de se limiter à la représentation visuelle directe d'une scène, pose la question suivante : "Quelle combinaison de capteurs, de traitement et de rendu donne la meilleure restitution possible ?". Par exemple, pour aider les patients à s'orienter dans les zones intérieures encombrées, ils ont mis au point un mode *TR* qui rend les zones de sol vide sous forme de phosphènes. Les FIGURES 1.24[9] et 1.25 illustrent respectivement le module externe et interne du Gennaris.



FIGURE 1.24 – Neuroprothèse GENNARIS - Composants externes

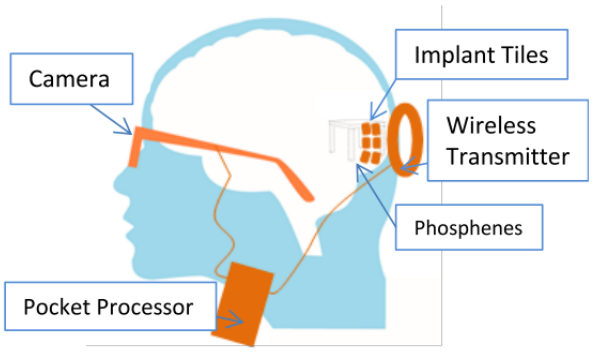


FIGURE 1.25 – Neuroprothèse GENNARIS - Composants internes [Lowery et al., 2015]

Bien que les essais cliniques de Gennaris n'aient pas encore commencé, la sécurité et l'efficacité du système ont été vérifiées de multiples façons [Wang et al., 2013a]. Les données préliminaires indiquent que les dommages causés au cortex sont minimes et que la transmission sans fil de l'énergie et des données entre la bobine externe et les tuiles implantées est possible.

9. https://www.monash.edu/mada/research/the-bionic-eye

ORION

Le dispositif Orion (Second Sight) hérite de nombreuses caractéristiques de l'Argus II, notamment les composants externes. Les composants implantables de l'Orion comprennent une bobine réceptrice et un circuit interne qui se place sur le crâne et contrôle les paramètres de stimulation délivrés par les électrodes. L'implant est composé de 60 électrodes (comme l'Argus II) placées sur le lobe occipital.

Le système Orion est la seule prothèse corticale visuelle dont les essais cliniques ont débuté. Une étude préliminaire a été réalisée sur une personne aveugle à l'aide d'un neurostimulateur. La cartographie des phosphènes et les tests de seuil ont été effectués pendant une période de 18 mois et aucune complication clinique n'a été constatée pendant et après l'implantation. Les phosphènes ont été déclenchés par une stimulation via tous les contacts et étaient dispersés sur le côté gauche du champ visuel. Après avoir confirmé la sécurité et la faisabilité de la stimulation de la surface corticale pour générer la perception des phosphènes, Orion a reçu l'approbation de la FDA pour les essais cliniques [Niketeghad and Pouratian, 2019]. La FIGURE 1.26 [10] illustre les éléments de la neuroprothèse ORION.

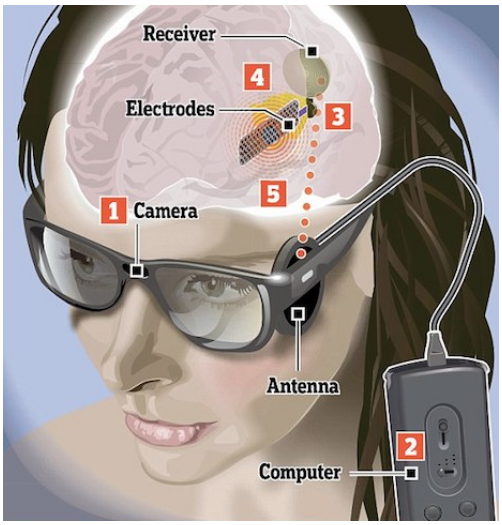


FIGURE 1.26 – Neuroprothèse ORION - Elle est composée de cinq éléments : (1) la caméra montée sur les lunettes, (2) le dispositif de calcul, (3) la matrice d'électrodes, (4) la bobine réceptrice et (5) la bobine émettrice.

Le dispositif a été implanté chez cinq personnes aveugles à l'Université de Californie à Los Angeles et au Baylor College of Medicine, et tous les sujets ont signalé la perception de phosphènes. Des tests comprenant la cartographie des phosphènes et des tests de seuil sont actuellement menés et l'intégration initiale de la caméra avec la stimulation est testée.

10. https://blog.adafruit.com/2016/11/01/bionic-eyes-from-second-sight/

UTAH ELECTRODE ARRAY et CORTIVIS

Initialement l'Utah Electrode Array (UEA) était un projet de l'université d'Utah et visait à construire un implant cortical de 100 électrodes (10x10). Des tests sur des chats et des études cliniques ont été menés et ont conduit à l'accréditation du FDA pour ce dispositif [Normann et al., 2009, Fernández et al., 2014]. La neuroprothèse visuelle corticale CORTIVIS [11] est un également un projet initié par l'université d'Utah visant à développer une prothèse visuelle intra-corticale [Fernández and Normann, 2017].

CORTIVIS tire parti du réseau d'électrodes de l'UEA.

Bien que les essais cliniques du projet CORTIVIS n'aient pas commencé, des études préliminaires ont été menées sur des patients épileptiques pour établir la sécurité de la procédure d'implantation ainsi que l'évaluation d'apparition des phosphènes. Tous les sujets ont perçu des phosphènes et ont toléré la procédure sans complications. La stimulation des aires visuelles a induit des phosphènes avec une conservation de la cartographie rétinotopique. Les phosphènes ont été décrits comme des lumières et des étoiles clignotantes, colorées ou non, et leur taille variait d'une "tête d'épingle" à l'occupation de la quasi-totalité du champ visuel, selon l'emplacement et les paramètres de stimulation.

Nous avons abordé les neuroprothèses corticales, les sous-sections suivantes traiteront des neuroprothèses se situant sur le corps géniculé latéral (CGL) et le nerf optique.

1.3.3 Neuroprothèses dans le corps géniculé latéral

Le corps géniculé latéral (CGL) reçoit l'information directement de la rétine et envoie des projections dans le cortex visuel primaire. Environ 90% des axones des ganglions rétiniens se connectent au noyau géniculé latéral dorsal. Une prothèse placée sur le CGL présente un potentiel de réadaptation pour les personnes atteintes de pathologies de la rétine ou du nerf optique. Comme pour les prothèses de la rétine, le moment de l'intervention sera une considération clé pour atténuer la dégradation neuronale en aval. Bien que les prothèses ciblant le CGL n'aient pas encore été implantées chez l'homme, les modèles animaux démontrent l'efficacité du dispositif par des réponses corticales [Panetsos et al., 2011] et une résolution grossière par des tâches de localisation d'objets [Pezaris and Eskandar, 2009].

1.3.4 Neuroprothèses sur le nerf optique

VERAART

Cet implant a vu le jour entre la Belgique et l'Allemagne. Il est constitué d'une électrode sous forme de spirale capable d'activer plusieurs zones du nerf optique. Des tests cliniques ont été réalisés sur une patiente de 59 ans atteinte de rétinite pigmentaire (RP)

11. https://www.neuraviper.eu/, https://entrain-vision.eu/

et aveugle depuis 2 ans. Au bout de 4 mois, les recherches permettaient de répertorier 1465 phosphènes répartis sur tous le champ de vision. Une étude deux ans après en 2000 a mis en évidence la difficulté de choisir à quels endroits les phosphènes apparaîtront suite à une stimulation des différentes zones. Les paramètres d'évolution sont l'amplitude de la stimulation, le nombre d'impulsions envoyés et la fréquence. Dans la dernière étude réalisée en 2005, le nombre de phosphènes perceptibles était de 109 [Veraart et al., 1998, Delbeke et al., 2003, Bradley et al., 2005]. Bien que la génération sélective de phosphènes soit plutôt rudimentaire, un patient implanté a réussi à localiser, discriminer et saisir de petits objets spécifiques [Duret et al., 2006].

CHINESE PROJECT FOR SIGHT (C-SIGHT)

Le dispositif C-Sight est composé d'une micro-caméra, d'une matrice de 13 électrodes et d'un processeur. C'est un des rares dispositifs qui prévoit l'implantation d'une caméra directement dans l'œil. Ce mécanisme permettrait de résoudre le problème d'asservissement du regard à la tête. Ainsi, un patient implanté pourrait observer son environnement en bougeant ses yeux et non plus en orientant sa tête dans la direction souhaitée [Chai et al., 2008]. La FIGURE 1.27 illustre les éléments de la neuroprothèse C-Sight.

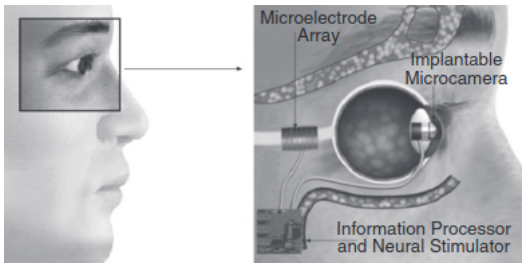


FIGURE 1.27 – Neuroprothèse C-SIGHT - Elle est composée de trois éléments, tous à l'intérieur du corps. On retrouve la caméra placée dans l'œil, le dispositif de calcul sous la peau et la matrice d'électrode autour du nerf optique.

1.3.5 Tableau récapitulatif des neuroprothèses présentées

Le Tableau 1.1 reprend la liste des neuroprothèses présentées jusqu'à maintenant. Il indique le nombre d'électrodes (ou photo-diodes) présent sur les implants, la présence d'essais cliniques et la commercialisation de ces neuroprothèses.

Dans la section 1.3, nous avons présenté plusieurs types de neuroprothèses visuelles. Une fois le dispositif implanté sur un non-voyant, il faut encore stimuler les différentes électrodes afin de faire apparaître des phosphènes dans le champ de vision. Il est cependant, encore aujourd'hui, très difficile de qualifier précisément la cartographie rétinotopique et la perception des phosphènes observés par les non-voyants implantés. Les quelques travaux

TABLEAU 1.1 – Tableau récapitulatif des implants présentés dans les parties précédentes. Les implants en jaune-orangé correspondent aux implants ayant été commercialisés.
* : Ce ne sont pas des électrodes, mais des micro-photodiodes.
EPI-R : implant épirétinien, SUB-R : implant sous-rétinien, SUPRA-C : implant suprachoroïdien

Type	**Nom**	**Nombre d'électrodes**	**Essais cliniques**	**Commercialisé ?**
EPI-R	ARGUS II (1.3.1)	60	Oui	Oui
EPI-R	IRIS II (1.3.1)	150	Oui	Non
EPI-R	EPI-RET3 (1.3.1)	24	Oui	Non
SUB-R	BRIP (1.3.1)	100	Oui	Non
SUB-R	ASR (1.3.1)	5000*	Oui	Non
SUB-R	ALPHA IMS (1.3.1)	1500*	Oui	Non
SUB-R	SYSTEME PRIMA (1.3.1)	378	Oui	Oui
SUPRA-C	BVA (1.3.1)	24	Oui	Non
SUPRA-C	STS (1.3.1)	49	Non	Non
Corticale	GENNARIS (1.3.2)	43	Non	Non
Corticale	ORION (1.3.2)	60	Oui	Non
Corticale	UEA et CORTIVIS (1.3.2)	100	Non	Non
Nerf optique	VERAART (1.3.4)	/	Oui	Non
Nerf optique	C-SIGHT (1.3.4)	13	Non	Non

qui visent à établir les schémas de perception décrivent une géométrie des phosphènes très variable (ronds, ellipses, traits) et une cartographie aléatoire. Les stratégies de stimulation (méthode de rendu) sont peu détaillées par les concepteurs des implants, mais elles semblent se rapprocher pour la plupart de la première méthode qui sera présentée dans le chapitre 2, traitant de notre simulateur de vision prothétique. Par la suite, nous allons détailler et expliquer quels sont les problèmes qui peuvent apparaître lors de l'utilisation de neuroprothèses. Nous présenterons ensuite quelques simulateurs de vision prothétique présents dans la littérature depuis 1992.

1.4 Simulateurs de vision prothétique

Un des enjeux majeurs de l'étude de la vision prothétique est la validation des méthodes de rendu. En effet, l'accès à des sujets de test non-voyants est très compliqué pour deux raisons : le faible nombre de patients implantés ; et la protection de ces derniers par les sociétés développant des implants. Cette difficulté pousse la majorité des chercheurs de ce domaine à élaborer des simulateurs de vision prothétique afin de tester des méthodologies de rendu ou d'évaluer des comportements et prises de décisions de la part de sujets voyants placés en conditions expérimentales. Dans un premier temps, nous aborderons les

problématiques liées à l'utilisation, et par conséquent, à la simulation de neuroprothèses visuelles (1.4.1). Ensuite, nous présenterons plusieurs simulateurs de vision prothétique, plus ou moins proches de la réalité physique (1.4.2).

1.4.1 Problèmes liés à l'utilisation de neuroprothèses visuelles

Bien que de nombreux progrès aient été réalisés dans le développement et la fabrication de neuroprothèses visuelles, de nombreux problèmes sont encore présents. Que ce soit des difficultés issues de l'implantation, ou des difficultés liées à la stimulation des tissus organiques, elles sont présentes peu importe la neuroprothèse visuelle. Nous détaillerons ici quelques-unes des difficultés majeures rencontrées lors de l'utilisation de neuroprothèses visuelles.

Non-conservation de la cartographie rétinotopique

Lors du développement d'une neuroprothèse visuelle, et plus particulièrement d'un implant, on s'attend à ce que la stimulation d'un ensemble d'électrodes génère dans le champ de vision du patient implanté un schéma relativement identique à celui de la stimulation initiale. En réalité, il existe une déformation de cette carte de stimulation, et cela peu importe l'implant considéré. Ce résultat a déjà été mis en évidence sur des chats en 1978 par Tusa et al. [Tusa et al., 1978]. Une étude réalisée sur l'implant Utah Electrode Array a montré des résultats similaires que l'on peut retrouver sur la Figure 1.28.

Phénomène d'évanescence

L'évanescence des phosphènes est un phénomène qui se produit lors des stimulations répétées d'une même micro-électrode. En effet [Schmidt et al., 1996] ont montré qu'une stimulation répétée d'une même électrode pendant un temps suffisamment long produit une diminution de l'intensité relative perçue par un sujet dans un contexte expérimental. Ce phénomène est lié au fonctionnement biologique des tissus récepteurs (rétine, cortex). La Figure 1.29 montre la diminution de réponse à des stimulis répétés.

Phénomène de persistance

La majorité des neuroprothèses visuelles souffrent de ce qu'on appelle la persistance phosphénique. Ce phénomène se produit dès lors qu'une électrode est stimulée. Dans un cas idéal, on souhaiterait qu'une stimulation phosphénique possède une réponse de type Dirac, ou on/off. En pratique, les phosphènes restent perceptibles dans le champ de vision quelques secondes après l'arrêt de la stimulation de l'électrode. Ce phénomène a notamment été mis en évidence sur l'Argus II de Second Sight [Brindley and Lewin, 1968, Echenique and Graffigna, 2011, Fornos et al., 2012].

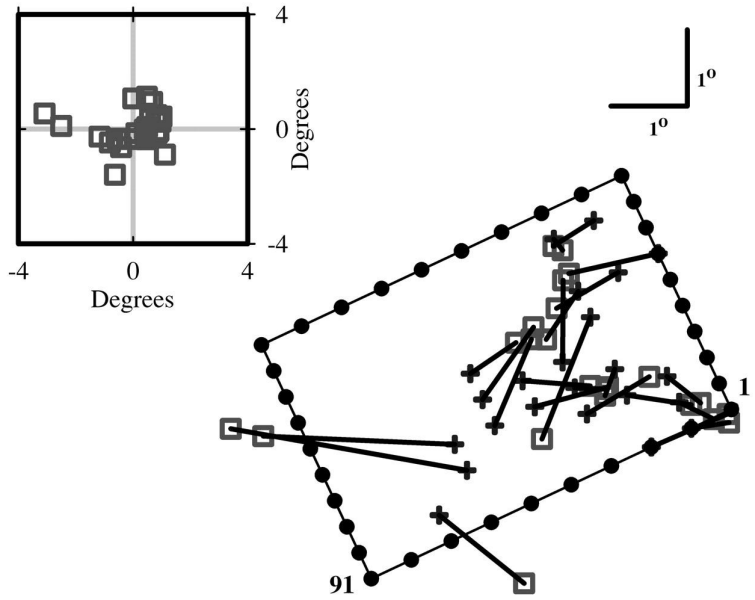


FIGURE 1.28 – On retrouve dans cette figure, pour chaque électrode simulée, la position des phosphènes générés dans le champ visuel du chat. L'unité de mesure est le degré de champ de vision. On constate que deux implants proches ne génèrent pas nécessairement de phosphènes proches dans le champ de vision. Les électrodes sont situées au niveau de la rétine. [Maynard, 2001]

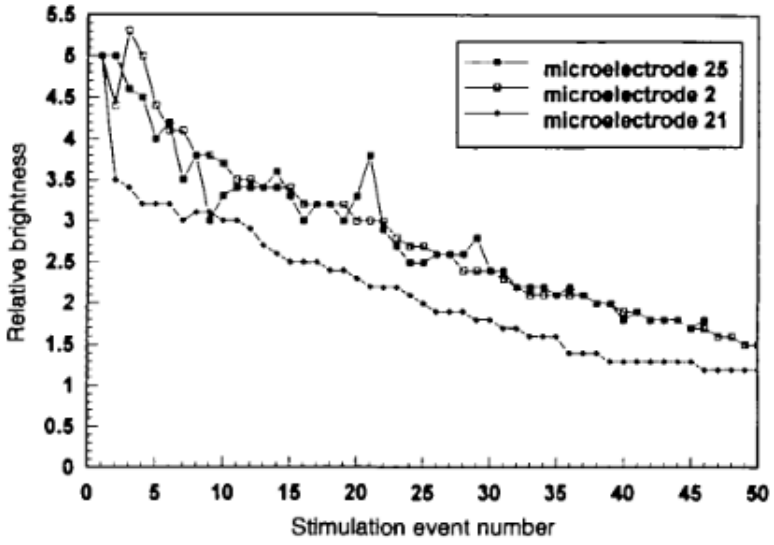


FIGURE 1.29 – Illustration du phénomène d'évanescence. Cette figure montre la perception de la luminosité relative des phosphènes produite par stimulation répétée, à 4 secondes d'intervalle, de trois micro-électrodes différentes. La micro-électrode 25 a été stimulée 46 fois, la micro-électrode 2 a été stimulée 50 fois et la micro-électrode 21 a été stimulée 49 fois. Extrait de (Schmidt et al., 1996). [Schmidt et al., 1996]

Résolution limitée

Cette difficulté est l'une des plus importantes à laquelle nous pouvons nous confronter. Les implants ayant passé les phases d'essais cliniques à l'heure actuelle ont une résolution extrêmement faible, de l'ordre de deux ou trois centaines d'électrodes pour les plus grands.

Pour donner un ordre d'idée de la perte d'information que cela représente, la Figure 1.30 montre une photographie d'un panda avant et après passage à la résolution de l'implant (dans ce cas la résolution la plus grande à l'heure actuelle, à savoir 21x18 = 378 électrodes). À titre de comparaison, la résolution de l'œil humain fonctionnel est estimée à 576 mégapixels (soit 1 million et demi fois plus que l'implant).

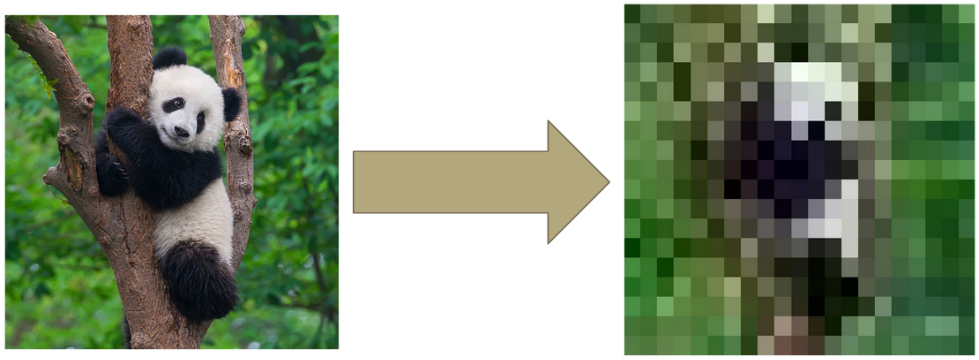


Figure 1.30 – À gauche une image de panda à la résolution 612x612 (environ 380 000 pixels), à droite la même image sous-résolue sur 21x18 = 378 pixels. La perte d'information est considérable.

Incapacité à contrôler la couleur, la forme, la taille et l'intensité perçue des phosphènes

De nombreuses formes de phosphènes ont été décrites lors de toutes les expérimentations, études et essais cliniques réalisés. Dans les descriptions des sujets, on retrouve les formes suivantes : des cercles, des carrés, des étoiles, des longues bandes rectilignes, des bandes courbes. La taille perçue est également variable par électrode pour un même sujet. La majorité des sujets perçoivent des tâches lumineuses blanches, mais il est arrivé dans quelques études passées que les sujets déclarent percevoir des phosphènes colorés allant du jaune au bleu [Hornig et al., 2007]. De plus, pour une même stimulation (en termes d'intensité), deux phosphènes ne "brilleront" pas de la même manière.

Apparition de plusieurs phosphènes sur une stimulation d'une seule électrode

Enfin, il arrive parfois que sur la stimulation d'une électrode, le courant électrique qui passe dans cette dernière stimule également les électrodes voisines. Ce phénomène peut conduire à l'apparition de plusieurs phosphènes, comportement qui n'est pas désiré. Ce phénomène est aussi l'une des raisons pour lesquelles les implants ne peuvent pas être

aussi petits que l'on souhaiterait. La densité d'électrodes sur l'implant est donc bornée.

Toutes ces problématiques se posent en conditions d'implantation réelle. Il faut par conséquent être capable de les modéliser lors de l'utilisation de simulateurs de vision prothétique. Dans la sous-section suivante, nous détaillons le fonctionnement de plusieurs simulateurs de l'état de l'art, depuis 1992 jusqu'en 2022.

1.4.2 Quelques simulateurs de la littérature

Au cours des 30 dernières années, de nombreux simulateurs ont été développés [Pissaloux et al., 2008, Tatur, 2011, Tatur, 2018]. Que ce soit pour des tâches d'évaluation de l'acuité visuelle, de lecture, de reconnaissance et localisation de visages ou d'objets, ou encore de navigation, chacun des simulateurs développés présente des caractéristiques qui lui sont propres. Nous détaillerons par la suite quelques-uns des simulateurs les plus pertinents dans un contexte de navigation et de reconnaissance d'objets pour pouvoir les comparer avec le simulateur que nous avons nous-même développé et qui sera présenté dans le chapitre 2.

Simulateur de Cha et al. – 1992

Le simulateur de vision prothétique proposé par [Cha et al., 1992a] est le premier à voir le jour. C'est un système de vision portable qui comporte une caméra et un écran pour faire la restitution. La simulation des phosphènes se fait par l'intermédiaire d'une carte perforée à différentes résolutions. C'est via l'utilisation de ce simulateur qu'ils ont montré que 625 électrodes sont suffisantes pour obtenir une acuité visuelle de 20/30. Ils ont également pu tester leur simulateur dans des tâches de lecture et de mobilité. Dans les tâches de lecture, la vitesse de lecture atteignait jusqu'à 100 mots par minute. En ce qui concerne les tâches de mobilité, les sujets devaient trouver leur chemin dans une succession de couloirs. Les sols étaient blancs et les murs noirs, et des obstacles blancs pendaient du plafond. L'objectif des sujets était d'atteindre la fin du parcours en évitant au maximum les obstacles [Cha et al., 1992b]. Les FIGURES 1.31 , 1.32 et 1.33 illustrent respectivement une tâche d'ouverture de porte, des exemples de différentes résolutions et un parcours de mobilité utilisé dans le cadre des expérimentations de Cha et al.

Cette expérimentation est historique, car c'est la première à utiliser un simulateur complet. On peut toutefois préciser que plusieurs problèmes liés à la simulation ne sont pas pris en compte dans ces études, notamment le nombre de niveaux de gris non limité et maximal (256).

Simulateur de Dagnelie et al. – 2001 / 2006 / 2007

Dagnelie et ses équipes ont développé plusieurs versions de simulateur correspondant aux différentes expérimentations qu'ils ont conduites. En 2001, ils construisent une première version de simulateur afin d'évaluer l'impact de la modification des hyperparamètres

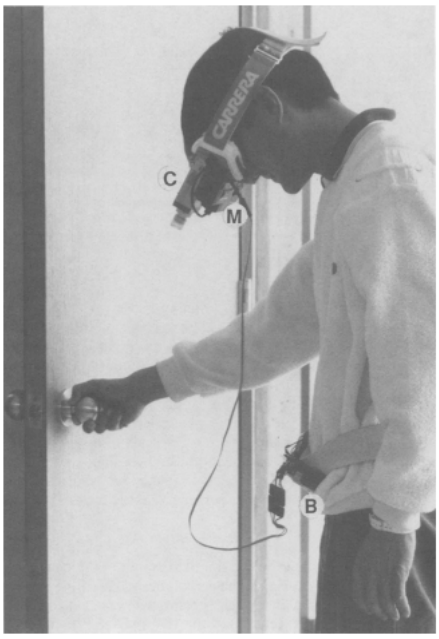


FIGURE 1.31 – Image du simulateur proposé par Cha et al. en 1992. La caméra (C) envoie le flux vidéo sur un écran monté dans les lunettes. La carte perforée simule la génération des phosphènes (M). Le système est alimenté en énergie par une batterie (B) [Cha et al., 1992a].



FIGURE 1.32 – Une image de base à gauche vue à travers deux cartes perforées respectivement de résolution 32x32 pour l'image du milieu et 16x16 pour l'image de droite [Cha et al., 1992a].

des implants dans des tâches de reconnaissance de visage, de lecture et de manipulation d'objets (réels et virtuels). Leur simulateur permet de faire varier la taille des phosphènes, les écarts entre deux phosphènes voisins, la résolution de l'implant, le taux d'électrodes brisées et le nombre de niveaux de gris [Dagnelie et al., 2001].

En 2006, ils proposent une nouvelle version de simulateur dans un contexte de placement de pions sur un jeu de dames. Cette nouvelle version inclut un asservissement du rendu phosphénique au mouvement des yeux. Ainsi, un mouvement oculaire d'observation se traduit en un déplacement de l'image restituée [Dagnelie et al., 2006].

Enfin, en 2007, ils proposent une version du simulateur dans un contexte de navigation [Dagnelie et al., 2007]. Leur expérience était découpée en deux parties. La première partie

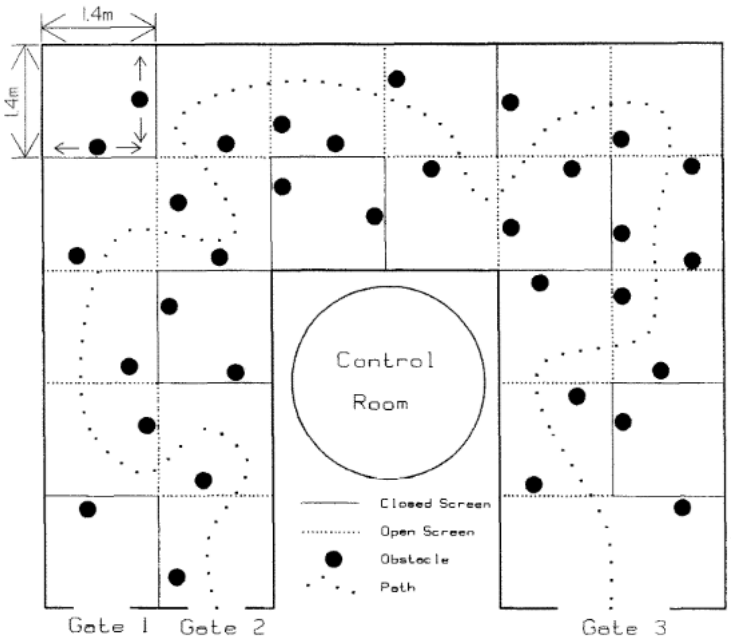


FIGURE 1.33 – Exemple de parcours à réaliser dans l'étude de navigation conduite par Cha et al. en 1992. Les points noirs sont des obstacles [Cha et al., 1992b].

consistait en un déplacement dans un bâtiment inconnu. Le but de la manipulation était de déterminer les paramètres de rendu minimaux pour réaliser un trajet simple et éviter des obstacles. Les sujets suivaient le parcours grâce à des indications orales. Trois tailles de matrices d'électrodes étaient évaluées lors de cette expérience (4x4, 6x10 et 16x16). Sans grande surprise, ils ont déterminé que la résolution nécessaire pour réussir la tâche est d'au moins 256 phosphènes (16x16). La deuxième expérimentation consistait en une navigation dans un environnement virtuel. Les sujets se déplaçaient à l'aide d'une manette et le rendu phosphénique était fait dans un casque. L'objectif était ici de réaliser un parcours le plus vite possible à travers différentes salles dont les murs étaient entièrement blancs et avec le sol et le plafond en noir. Aucun obstacle n'était cependant présent dans cette deuxième expérimentation. Les FIGURES 1.34 et 1.35 illustrent respectivement une tâche de navigation en conditions réelles et une tâche de navigation en environnement virtuel simulé.

Simulateur de Denis et al. – 2012

Le simulateur développé par l'équipe de [Denis, 2014], a été utilisé pour évaluer ce qu'ils appellent l'*approche par localisation*. Cette méthode consiste à construire une méthode de restitution qui va dans un premier temps donner un indice sur l'endroit où se trouve l'objet à saisir en allumant uniquement des phosphènes vers l'endroit où se trouve l'objet dans le champ de vision et en ne stimulant pas les autres. Trois tailles de matrices différentes ont été utilisées dans cette étude (6x10, 15x18, 32x38). Cette approche a permis d'avoir de



FIGURE 1.34 – Image de la première expérimentation de Dagnelie et al. À gauche, l'environnement réel et, à droite, l'environnement restitué sous forme de phosphènes [Dagnelie et al., 2007].

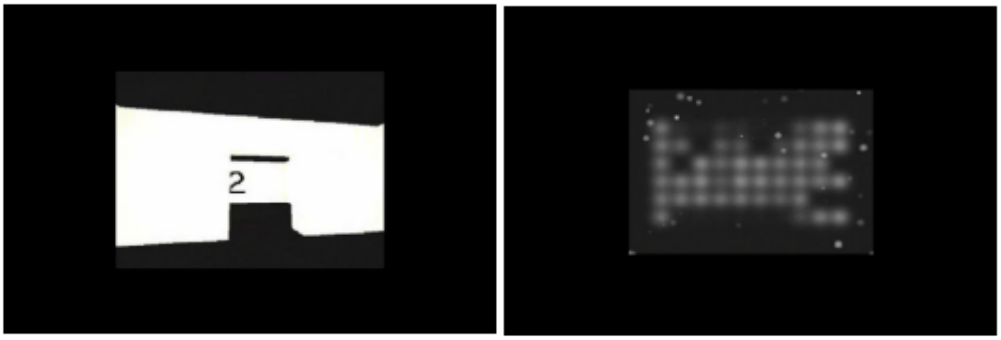


FIGURE 1.35 – Image de la seconde expérimentation de Dagnelie et al. À gauche, l'environnement virtuel et, à droite, l'environnement virtuel restitué sous forme de phosphènes [Dagnelie et al., 2007].

meilleures performances que les rendus classiques avec une matrice de résolution 15x18. La FIGURE 1.36 illustre le rendu phosphénique de divers objets (prise, café, agrafeuse) avec différentes résolutions (6x10, 15x18 et 32x38). Ces rendus ont été utilisés dans une tâche de préhension d'objets.

Ce simulateur a été repris dans une autre expérimentation en 2015 [Macé et al., 2015] avec quatorze sujets.

Simulateur de Mc Carthy et al. - 2015

En 2015, Mc Carthy et al. proposent un rendu qualifié de rendu de "*distance augmentée*" [McCarthy et al., 2014]. Le modèle d'implant simulé est composé de 30x35 électrodes alignées sur une grille rectangulaire. Ils utilisent un algorithme qui permet de différencier les murs, sols et plafond et par-dessus un encodage de la distance des éléments grâce à une variation de l'intensité lumineuse. En plus de cela, les obstacles sont mis en évidence. Les 8 sujets impliqués dans l'expérience devaient franchir un couloir encombré sans heurter les obstacles le plus rapidement possible. Les résultats montrent que le rendu "*distance*

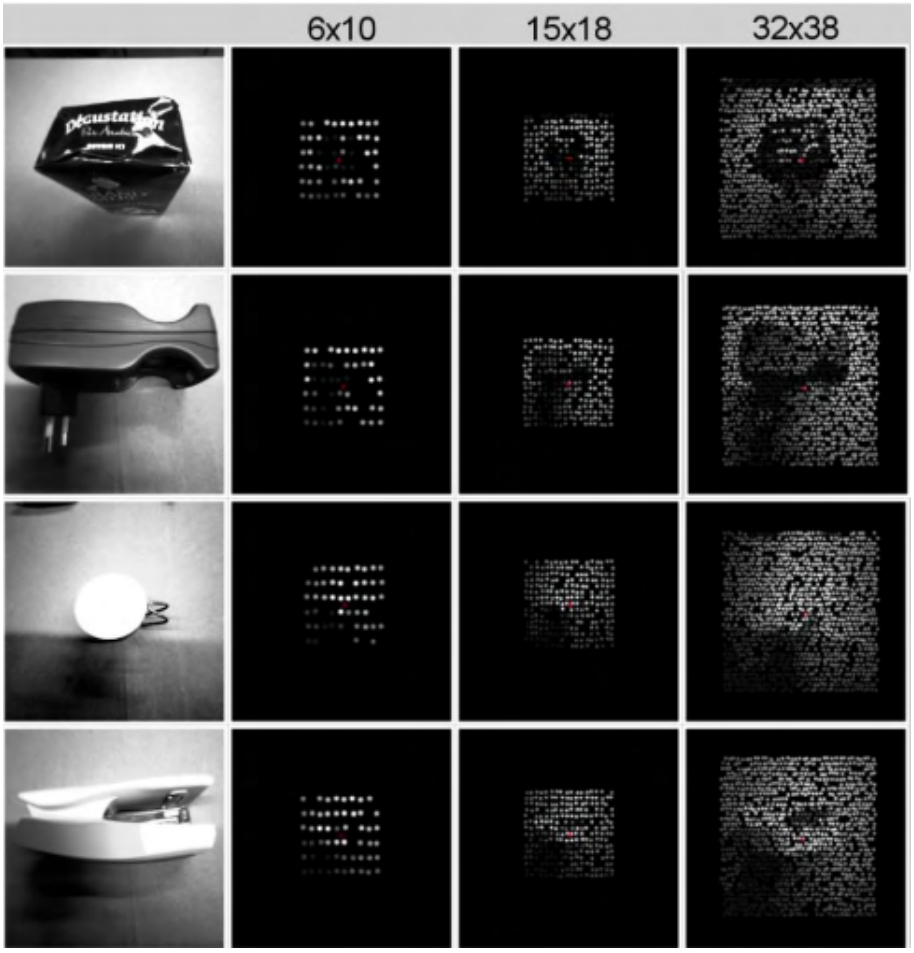


FIGURE 1.36 – Rendus utilisés dans une tâche de préhension d'objets conduite par [Denis, 2014]. Différentes résolutions de rendus phosphéniques sont présentées (6x10, 15x18, 32x38). Les objets montrés sont un paquet de café, une prise électrique, une tasse de café et une agrafeuse.

augmenté" limite le nombre de collisions avec les obstacles. Les FIGURES 1.37 et 1.38 illustrent respectivement les rendus crées et utilisés dans les expériences réalisées par Mc Carthy et al. ainsi qu'un sujet en expérimentation.

Simulateur de Zapf et al. - 2014 / 2015

Le simulateur développé par l'équipe de Zapf a été utilisé dans deux expérimentations de navigation [Zapf et al., 2014]. La première étude était divisée en deux tâches : une tâche consistant en l'évitement d'obstacles dans un couloir virtuel, et une tâche de suivi de ligne blanche au sol pour éviter des voitures statiques. Les résolutions des matrices utilisées sont de 10x10 et 20x20 électrodes simulées. Les sujets utilisaient le clavier et la souris ou une manette pour naviguer dans l'environnement 3D. C'est une des rares expériences où un effet d'apprentissage a été obtenu (quelle que soit la technique d'interaction, la vitesse

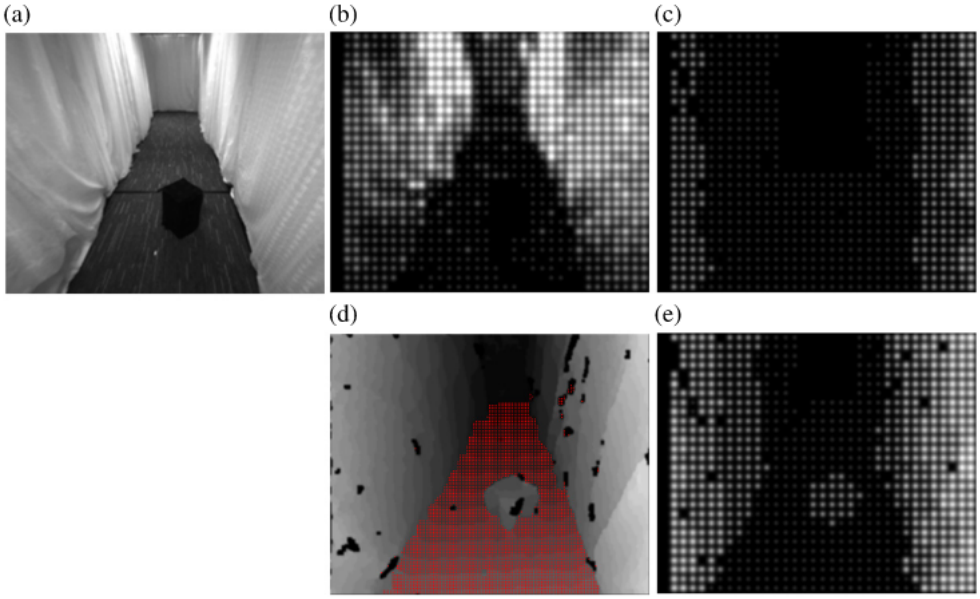


FIGURE 1.37 – Rendus phosphéniques développés dans l'étude de Mc Carthy et al. [McCarthy et al., 2014]. (a) Photo de départ en nuances de gris. (b) Rendu de distance, (c) et (d) Extraction de murs et d'objets. (e) Rendu de distance augmentée.

moyenne des sujets augmentait au fur et à mesure des essais).

Les FIGURES 1.39 et 1.40 illustrent les rendus à différentes résolutions, ainsi qu'un parcours à effectuer, mis en place dans les expériences réalisées par Zapf et al. en 2014.

La seconde expérimentation portait sur l'ajout de la possibilité pour des personnes ayant une vision tubulaire d'accéder à des informations de périphérie [Zapf et al., 2015]. En plus des deux manipulations précédentes, les sujets devaient se déplacer en évitant des piétons statiques et en mouvement. Toutes les interactions étaient réparties le long d'un parcours. Pour chacune de ces tâches, l'ajout d'information de périphérie grâce aux phosphènes a permis d'améliorer les performances des sujets.

Les FIGURES 1.41 et 1.42 illustrent le rendu tubulaire, ainsi qu'un parcours à effectuer, mis en place dans les expériences réalisées par Zapf et al. en 2015.

Simulateur de Vergnieux et al. - 2017

Ce simulateur a été développé pour évaluer et comparer différentes méthodologies de rendu dans un contexte de navigation [Vergnieux et al., 2014]. Dans leur étude, les auteurs comparent quatre méthodologies de rendu : le rendu *Irrlicht* qui correspond simplement au rendu issu du moteur graphique du même nom, le rendu *VPS-LimInf* qui réalise un traitement standard *scoreboard* sans limitation de distance du champ de vision, le rendu VPS-Distance qui limite la vision en termes de profondeur et le rendu *VPS-FilDeFer* qui met en évidence les informations structurelles de la scène. La FIGURE 1.43 montre les

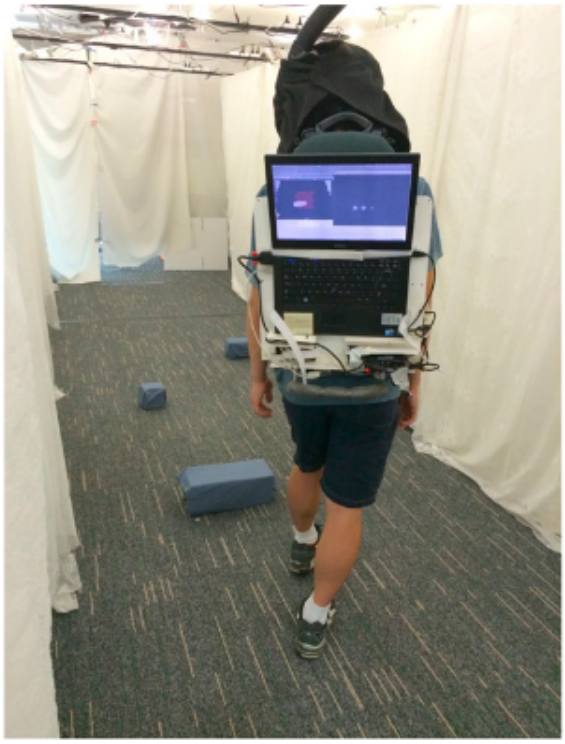


FIGURE 1.38 – Image prise pendant l'expérimentation de Mc Carthy en 2015. [McCarthy et al., 2014]. Il s'agit d'une tâche de navigation dans laquelle il faut éviter des obstacles placés sur le sol.

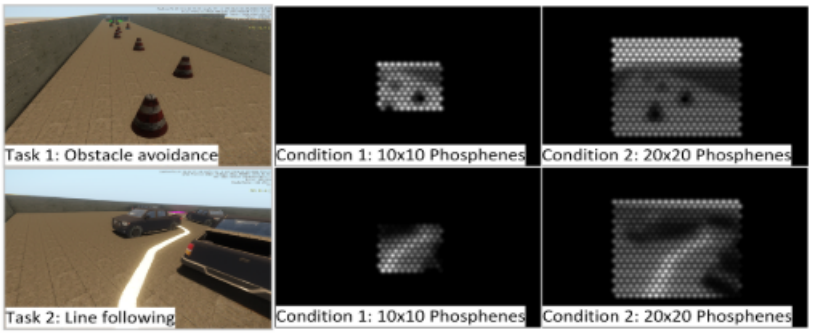


FIGURE 1.39 – Rendus phosphéniques utilisés dans l'étude de Zapf et al. [Zapf et al., 2014]. Les résolutions montrées sont 10x10 et 20x20. On retrouve deux tâches de navigation : une tâche d'esquive d'objets et une tâche de suivi de ligne blanche au sol.

différents rendus modélisés dans le simulateur. Le but de la manipulation est de laisser les sujets se déplacer dans un environnement labyrinthique qui comporte des joyaux placés aléatoirement. Les sujets doivent ramener 9 joyaux au point de départ, le plus vite possible, sachant qu'ils ne peuvent en porter que trois à la fois. À l'issue de cette navigation, les sujets doivent tracer le plan en vue de dessus de l'environnement dans lequel ils avaient évolué. La

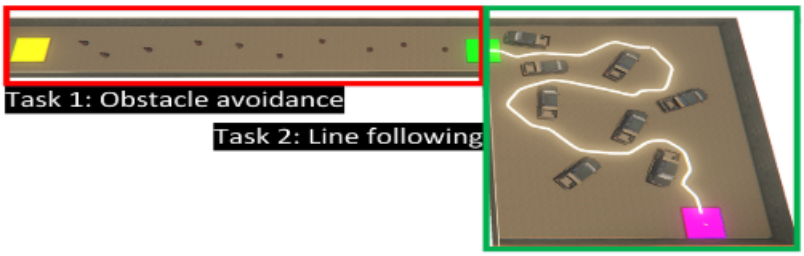


FIGURE 1.40 – Parcours à effectuer dans l'étude de Zapf et al. en 2014 [Zapf et al., 2014]. Le parcours est divisé en deux zones, une zone dans laquelle il faut esquiver des obstacles au sol, et une zone dans laquelle il faut suivre une ligne blanche au sol.



FIGURE 1.41 – Représentation de la vision tubulaire utilisée dans l'expérience de Zapf et al. en 2015 [Zapf et al., 2015].

FIGURE 1.44 montre des exemples de plans tracés par les sujets après la navigation. Grâce à ces informations, les auteurs évaluent à la fois la performance de navigation (vitesse et nombre d'objets ramenés) et la capacité des sujets à construire une carte mentale de l'endroit exploré. Ils mettent aussi en avant le fait que les sujets qualifiés de "joueurs" (jeux vidéo) réussissent mieux cette tâche.

Simulateur de Sanchez-Garcia - 2018 / 2019

Le simulateur proposé par Sanchez-Garcia a pour objectif d'évaluer et de comparer différents rendus construits sur la base de prédictions de réseaux de neurones [Sanchez-Garcia et al., 2018, Sanchez-Garcia et al., 2019]. La FIGURE 1.45 montre les différents rendus modélisés dans le simulateur. Les sujets devaient effectuer une tâche de classification de scènes d'intérieur en visualisant des images statiques et des séquences vidéo sous différentes restitutions. La tâche donnée aux sujets était de reconnaître un environnement intérieur parmi quatre types de pièces, et les types d'objets présents dans chaque scène. La sélection de scènes d'intérieur comporte des chambres, des cuisines, des salons et des salles à manger, et la sélection d'objets comporte d'éviers, de réfrigérateurs, de fours/micro-ondes, de tables, de chaises, de téléviseurs/ordinateurs portables, de lits et de canapés. Les réseaux de neurones sous-jacents sont issus de [Mallya and Lazebnik, 2015, He et al., 2017]. Le schéma de construction du rendu combiné que les auteurs proposent est illustré sur la FIGURE 1.46.

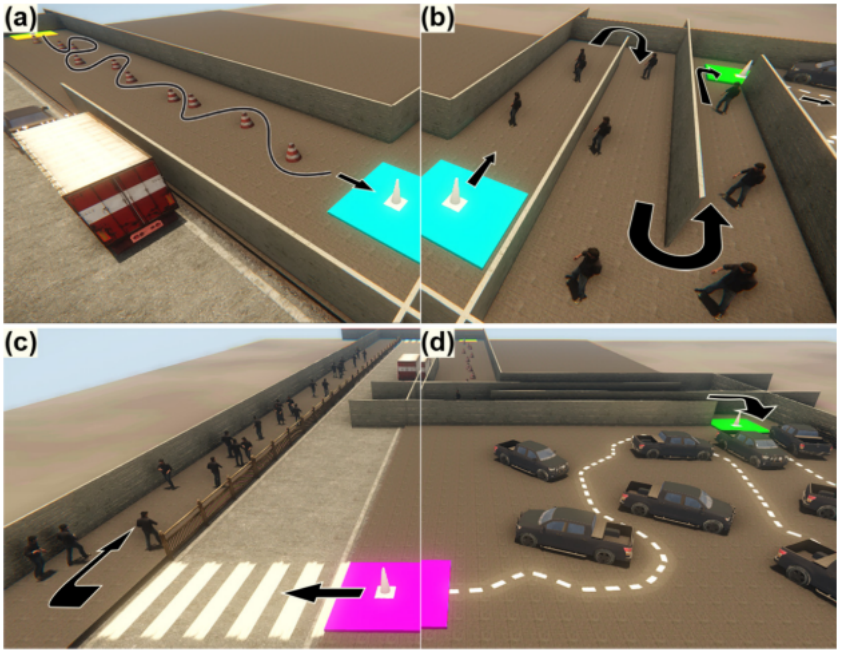


FIGURE 1.42 – Parcours à effectuer dans l'étude de Zapf et al. en 2015 [Zapf et al., 2015]. Le parcours est divisé en quatre zones : (a) une zone dans laquelle il faut esquiver des obstacles au sol, (b) une zone dans laquelle il faut esquiver des personnes immobiles, (c) une zone dans laquelle il faut suivre une ligne blanche au sol et enfin (d) une zone dans laquelle il faut éviter des personnes mobiles.

Tous les sujets ont testé les images avant les vidéos. La résolution de l'implant simulé est de 32x32. La première tâche est composée de huit images statiques et de quatre vidéos de différents environnements intérieurs. La deuxième tâche consiste à identifier le type de pièce auquel correspond chaque exemple, c'est-à-dire que les sujets doivent déterminer les types de pièces en fonction de ce qu'ils voient sur chaque exemple phosphénique. Douze sujets se sont portés volontaires pour cette expérience. Les sujets avaient entre 20 et 60 ans.

Leurs résultats montrent que les algorithmes d'apprentissage profond peuvent mieux utiliser la résolution limitée en mettant en évidence les caractéristiques saillantes telles que les bords structurels et la segmentation d'objet pour la vision prothétique simulée. Leur nouvelle approche a une bonne performance dans la reconnaissance d'environnements intérieurs, même à basse résolution.

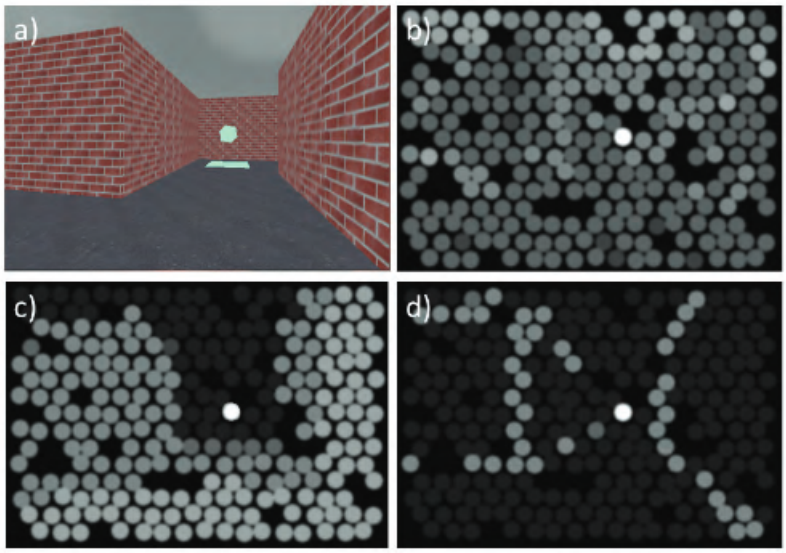


FIGURE 1.43 – Représentation phosphéniques modélisés et utilisés dans les études de Vergnieux et al. On retrouve : a) le rendu Irrlicht original (IRR-LimInf), b) le rendu brut classique (VPS-LimInf), c) le rendu prothétique basé sur la distance existant entre les éléments et la caméra (VPS-Distance) et d) le rendu VPS-FilDeFer.

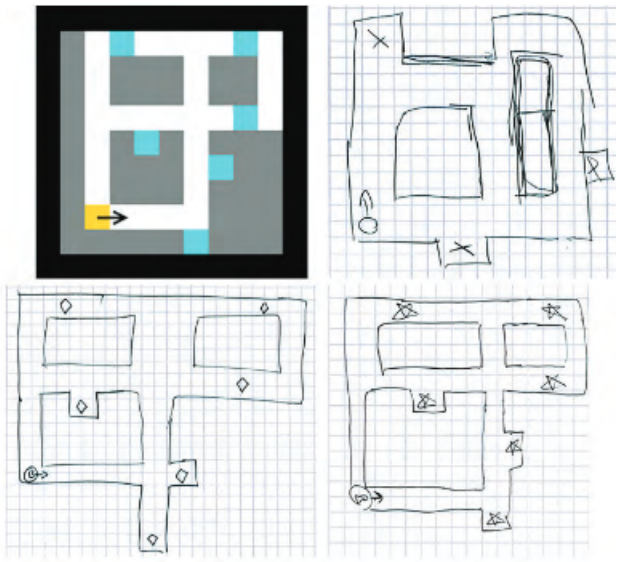


FIGURE 1.44 – Exemples de plans dessinées par les sujets. La carte de l'environnement est représentée en haut à gauche. Les trois autres images sont des représentations produites par des sujets après exploration.



FIGURE 1.45 – Exemples d'environnements intérieurs représentés avec un implant 32x32 dans le cadre de l'étude de Sanchez-Garcia. De gauche à droite, on retrouve systématiquement l'image issue de la simulation, une segmentation d'objet, une extraction de contour qui utilise un filtrage de Canny et enfin une prédiction réalisée par des réseaux de neurones pour extraire la structure et les objets de la scène.

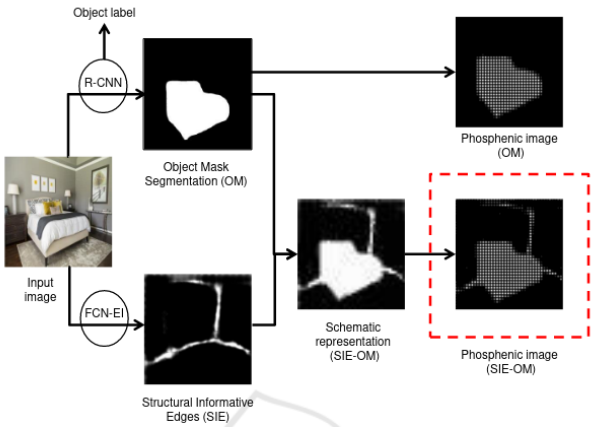


FIGURE 1.46 – La méthode de création du rendu réseau de neurones proposé par Sanchez-Garcia et al. À partir d'une image d'entrée, les bords informatifs structurels sont extraits à partir de [Mallya and Lazebnik, 2015] et la segmentation d'objets est générée à partir de [He et al., 2017]. Les informations extraites sont superposées pour créer l'image phosphénique.

1.4.3 Tableau récapitulatif des simulateurs recensés pour des tâches de navigation ou de manipulation/reconnaissance d'objets

Le tableau 1.2 regroupe plusieurs simulateurs conçus pour simuler des tâches de navigation ou de manipulation d'objets. Ceux marqués en jaune correspondent aux simulateurs

utilisant des matrices d'électrodes de résolution réaliste.

Notons que la tendance sur les quatre dernières années est de simuler des implants de résolution élevée (en vert sur le tableau), tout en basant la construction des rendus sur des méthodes de traitement d'image, notamment par l'utilisation de réseaux de neurones. Ceci est probablement basé sur l'hypothèse que la résolution des implants va augmenter avec le temps, hypothèse pas encore vérifiée aujourd'hui. C'est une des raisons pour lesquelles nous utiliserons une résolution plus faible que celles-ci (21x18) dans les expériences des chapitres 3 et 4.

La section suivante est consacrée à la présentation de techniques d'extraction de caractéristiques des images utilisés dans certains des simulateurs, sur lesquelles nous nous baserons pour construire nos rendus.

1.5 Apprentissage profond et réseaux de neurones

En 2012, le réseau de neurones AlexNet [Krizhevsky et al., 2012] remporte le concours de classification sur la base de données ImageNet, relançant ainsi l'intérêt pour ces modèles. L'amélioration de la puissance de calcul des cartes graphiques et l'accès de plus en plus facile a de larges bases de données a occasionné l'essor des réseaux de neurones les 10 années suivantes. Nous ferons dans cette section un bref historique des réseaux de neurones, les principales caractéristiques de ces derniers et le cadre applicatif que nous envisageons quant à leur utilisation.

1.5.1 Historique

C'est en 1943 que les réseaux de neurones apparaissent pour la première fois avec l'introduction par W. MCCulloch et W. Pitts des neurones formels [McCulloch and Pitts, 1943] qui sont des abstractions mathématiques des neurones physiologiques. Leur objectif était de prouver que le cerveau est équivalent à une machine de Turing, et que la pensée est un mécanisme purement physique et logique. Ils voient les neurones comme des fonctions qui transforment une valeur d'entrée en valeur de sortie (peu importe le type de valeur).

En 1949, Hebb formule une règle aujourd'hui connue sous le nom de "loi de Hebb" [Hebb, 1949]. Elle permet de préciser sous quelles conditions les poids doivent évoluer dans un réseau de neurones comme montré sur la Figure 1.47. La formulation est relativement simple : si deux neurones qui se suivent sont activés en même temps, alors la force de leur connexion augmente. Cette règle est toujours utilisée de nos jours.

En 1958, F. Rosenblatt développe le modèle du Perceptron [Rosenblatt, 1958]. C'est un

TABLEAU 1.2 – Tableau récapitulatif des simulateurs de vision prothétique utilisés dans des tâches de manipulation / reconnaissance d'objets (**MO/RO**), de reconnaissance de pièces d'intérieur (**RPI**) ou de navigation (**NAV**). En jaune-orangé, les simulateurs n'utilisant que des implants cliniquement réalistes. En vert, les résolutions des derniers simulateurs sont plus grandes que les résolutions des implants commercialisés à l'heure actuelle.

Publication	**Tâche**	**Résolution de l'implant**	**Phosphènes - (Niveaux de gris)**
[Cha et al., 1992a]	TEST	10x10, 16x16, 25x25, 32x32	Points
[Dagnelie et al., 2001]	MO	10x10, 16x16, 25x25, 32x32	Ronds - (2, 4, 6, 8)
[Boyle et al., 2002]	RO	10x10, 16x16, 25x25	Ronds - (2, 3)
[Hayes et al., 2003]	RO	4x4, 6x10, 16x16	Carrés, Ronds - (1 - 8)
[Dowling et al., 2005]	NAV	24x32	(8, 256)
[Dagnelie et al., 2006]	MO	6x10	Ronds, Gaussien
[Fornos et al., 2008]	MO	17920, 1991, 498, 221, 124	Carrés
[Wang et al., 2008]	NAV	6x10	Ronds, Gaussien, 256
[Zhao et al., 2010]	RO	8x8, 16x16, 24x24, 32x32, 48x48, 64x64	Carrés, Ronds
[Lieby et al., 2011]	NAV	30x30	/
[Lu et al., 2013]	RO	16x16, 24x24, 32x32	Ronds, Gaussien - (2, 4, 6, 8)
[Parikh et al., 2013]	RO	6x10	Ronds, Gaussien - (1 - 8)
[Denis et al., 2014]	MO	6x10, 15x18, 32x38	Ronds, Gaussien - (1 - 8)
[Hu et al., 2014]	RO	8x8, 16x16, 24x24, 32x32	Ronds, Gaussien - (2, 4, 6, 8)
[McCarthy et al., 2014]	NAV	30x35	Ronds, Gaussien - (1 - 8)
[Zapf et al., 2014]	NAV	10x10, 20x20	Ronds, Gaussien - 256
[Vergnieux et al., 2014]	MO	9, 10x10	Ronds
[Zapf et al., 2015]	NAV	10x10	Ronds, Gaussien - 256
[Vergnieux et al., 2017]	NAV	15x18	Ronds, Gaussien - (2, 4, 6, 8)
[Sanchez-Garcia et al., 2018]	RPI	32x32	Ronds, Gaussien - (1 - 8)
[Li et al., 2018]	RO	18x24, 42x56, 75x100, 117x156, 174x232	Ronds, Gaussien - (1 - 8)
[Sanchez-Garcia et al., 2019]	RPI	40x40	Ronds, Gaussien - (1 - 8)
[van Steveninck et al., 2022]	NAV	26x26	Ronds, Gaussien

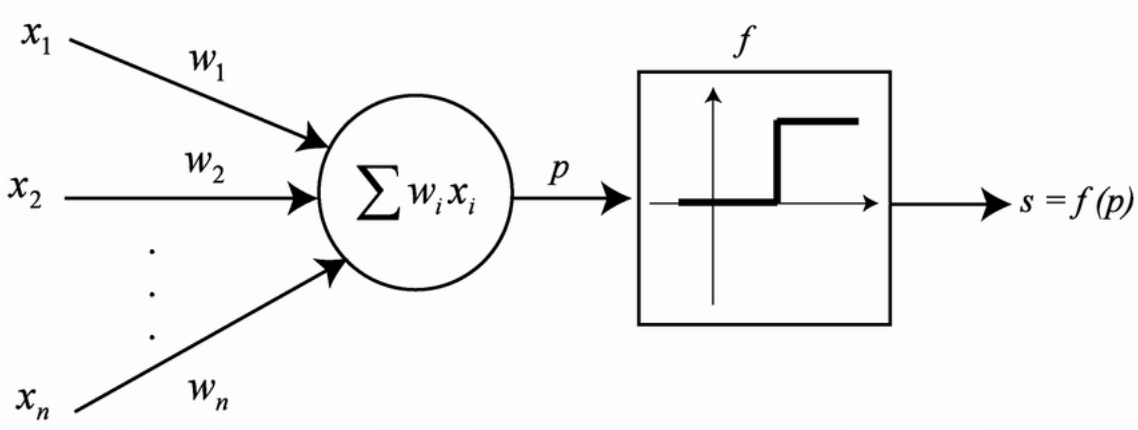


FIGURE 1.47 – Neurone formel. Les valeurs d'entrées sont les xi. Elles sont pondérées par les wi puis sommées et données en entrée de la fonction d'activation (ici un seuil), ce qui produit la valeur finale.

réseau de neurones inspiré du système visuel. Il possède deux couches de neurones : une couche de perception et une couche liée à la prise de décision. C'est le premier système artificiel capable d'apprendre par expérience. Il a été utilisé pour la première fois sur un problème de classification linéaire binaire de vecteur. En 1969, Minsky et Papert montrent que ce genre de réseaux (perceptrons simple) ne peut apprendre à construire que des fonctions solutions pour des problèmes linéairement séparables [Minsky and Papert, 1969]. En 1986, on assiste à l'apparition des premiers réseaux multi-couches qui permettent de résoudre désormais des problèmes non-linéaires (Rumelhart Hinton, 1986). C'est grâce à Yann LeCun et à la méthode de rétropropagation des gradients que ces avancées sont possibles [Cun, 1986]. En effet, la technique de la rétropropagation des gradients rend l'entraînement des réseaux de neurones multicouches beaucoup plus efficaces (en termes de calcul). Un des problèmes majeurs à cette époque reste cependant la quantité de temps colossale nécessaire pour entraîner un réseau.

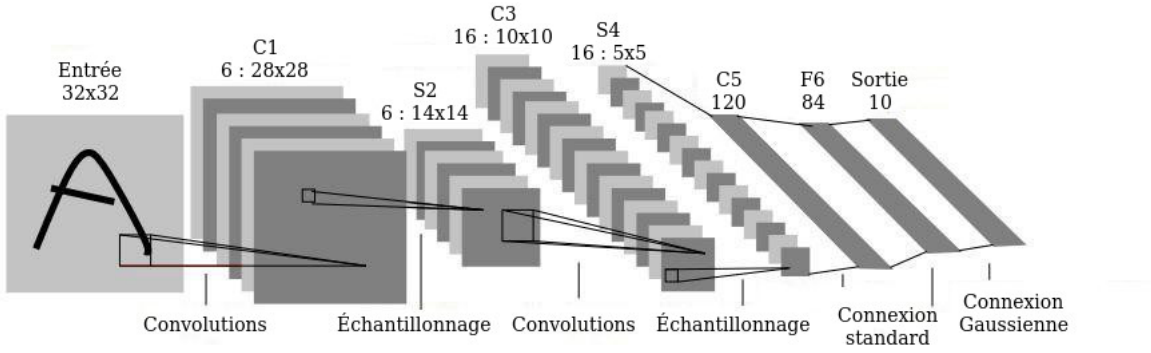


FIGURE 1.48 – Réseau convolutionnel LeNet5, proposé par LeCun en 1998 [LeCun et al., 1998].

Dans les années 2010, la puissance des cartes graphiques augmente de manière drastique, ce qui permet d'entraîner des réseaux de plus en plus performants et de manière rapide. C'est ainsi qu'on assiste à la naissance d'AlexNet [Krizhevsky et al., 2012], que l'on peut voir comme une mise à l'échelle de LeNet5 [LeCun et al., 1998], illustré sur la FIGURE 1.48, pour des problèmes plus complexes. D'autres réseaux convolutifs de plus en plus élaborés ont par la suite gagner le concours ImageNet : VGG [Simonyan and Zisserman, 2014] et ResNet [He et al., 2016].

Aujourd'hui, les outils de calcul surpuissants et la multitude de nouveaux réseaux engendrés par les recherches permettent de résoudre une large variété de problèmes tels que la classification, la segmentation sémantique, l'apprentissage par renforcement ou encore le traitement du langage naturel.

En parallèle, certains travaux s'attachent à créer des architectures plus compactes, et à compresser les réseaux, de façon à pouvoir les embarquer dans de petits systèmes (smartphones, tablettes, raspberry).

On ne développera pas de façon exhaustive ni le fonctionnement, ni la construction de tels réseaux, mais dans la suite, on présentera ceux qui nous sont utiles.

1.5.2 Nos besoins en termes de réseaux de neurones

Pour construire une représentation phosphénique pertinente, nous avons besoin d'extraire les caractéristiques importantes de la scène visuelle. Ces caractéristiques peuvent être par exemple les objets présents dans la scène, la structure des éléments présents ou encore le type de scène (intérieur ou extérieur).

Segmentation sémantique

L'objectif de la segmentation sémantique est de partitionner une image donnée en plusieurs zones visuellement significatives ou intéressantes pour l'analyse de l'image et la compréhension visuelle. Un exemple de segmentation sémantique issu de [Hao et al., 2020] est montré sur la FIGURE 1.49. En pratique, cela consiste à attribuer à chaque pixel d'une image une classe sémantique (type d'éléments : voiture, passage piéton, ...).

La segmentation joue un rôle central dans de nombreux domaines d'applications, notamment dans l'analyse d'images médicales (extraction des limites de tumeurs et mesure des volumes de tissus), les véhicules autonomes (détection des surfaces navigables et des piétons), la surveillance vidéo ou encore la réalité augmentée. De nombreux algorithmes de segmentation d'images ont été développés dans la littérature comme par exemple la méthode des contours actifs, ou bien les méthodes basées clustering ("*Mean-Shift*"), ou encore les méthodes de coupure de graphe. Cependant, au cours des dernières années, les modèles d'apprentissage profond ont donné naissance à une nouvelle génération de modèles de segmentation d'images dont les performances ont été remarquablement améliorées, at-



FIGURE 1.49 – Exemple de segmentation sémantique 2D : (a) image d'entrée, (b) prédiction. Chaque pixel de l'image d'entrée est associé à une classe d'objet. Dans l'image de sortie, la couleur des pixels est modifiée selon la classe de l'objet associé. Cette technique permet d'identifier le placement des objets [Hao et al., 2020].

teignant souvent des précisions largement plus élevées que les approches classiques sur les bases de données d'images habituellement utilisées (MSCOCO, PASCALVOC 2012).

Les réseaux de neurones convolutifs sont utilisés pour cette tâche depuis [Long et al., 2015]. Le problème est formulé comme une classification à chaque emplacement de pixel, et le réseau doit produire une carte de probabilités de présence pour chaque type d'objet.

Les architectures neuronales de segmentation d'image présentent majoritairement deux caractéristiques principales. L'architecture d'auto-encodeur est souvent au cœur du réseau ; elle consiste en un encodeur qui effectue une extraction de caractéristiques, accompagné d'un décodeur qui effectue le sur-échantillonnage de l'"espace latent" à partir des cartes de caractéristiques intermédiaires de l'encodeur. Plusieurs variantes de ce concept ont été proposées, la plus populaire étant U-Net [Ronneberger et al., 2015] pour la segmentation des lésions cutanées. Un autre exemple de réseaux de type encodeur-décodeur est proposé par Kendall et al. : SegNet [Kendall et al., 2015] pour la segmentation de scènes urbaines.

Un deuxième module commun dans l'architecture des réseaux de segmentation est la pyramide spatiale de caractéristiques, introduite pour combiner les caractéristiques de l'encodeur à différentes échelles afin de reconnaître plus efficacement les objets à différentes échelles. La mise en commun de pyramides spatiales (*Pyramid Spatial Pooling*) est un exemple de cette technique, dans laquelle plusieurs convolutions avec différentes tailles de noyau sont appliquées simultanément à la carte de caractéristiques pour extraire des informations multi-échelles. Cette idée est utilisée dans DeepLab V3+ [Chen et al., 2017], l'un des réseaux les plus performants, combiné à une architecture auto-encodeur. Certains des travaux les plus récents, comme NAS-FPN [Ghiasi et al., 2019], proposent d'apprendre à combiner de manière optimale plusieurs échelles de la carte de caractéristiques.

La FIGURE 1.50 illustre des exemples de prédictions réalisées avec un réseau encodeur-décodeur avec une base convolutive ResNet-50 [Baheti et al., 2019]. La FIGURE 1.51 illustre des exemples de prédictions réalisées avec le réseau DeepLab V3+ [Chen et al., 2017].



FIGURE 1.50 – Exemple de résultats de segmentation sémantique sur des images d'extérieur avec un réseau ResNet-50 comme encodeur. La première ligne montre les images d'entrée et la deuxième ligne montre la vérité terrain. La troisième rangée montre le résultat de la segmentation sémantique où différentes couleurs correspondent aux différentes classes [Baheti et al., 2019].



FIGURE 1.51 – Exemples de résultats de segmentation sémantique en utilisant le réseau DeepLab V3+. La première ligne montre les images d'entrée et la seconde ligne montre les prédictions réalisées par le réseau [Chen et al., 2017].

Extraction de structure

Le deuxième type d'informations que nous souhaitons pouvoir extraire pour construire nos rendus sont les informations structurelles des scènes visuelles. Dans le cas d'un environnement intérieur, il s'agit des jointures entre sols, murs et plafonds. Dans le cas d'un environnement extérieur, il s'agit des jointures entre les bâtiments, le sol et les trottoirs. Sur la FIGURE 1.52, des exemples d'extraction de structures sont proposés.

Ce genre de travaux est moins présent dans la littérature que ceux qui concernent la segmentation sémantique, mais ils ne sont pas absents pour autant. L'objectif de ces travaux est de pouvoir mettre en évidence les informations structurelles des scènes même si de l'encombrement est présent, comme cela peut être vu sur la FIGURE 1.52. Les réseaux utilisés dans ces travaux sont basés sur les mêmes techniques que celles présentées dans la partie relative à la segmentation sémantique. Le problème peut en effet être naturellement formulé comme une classification binaire de chaque pixel, en structure ou non-structure.

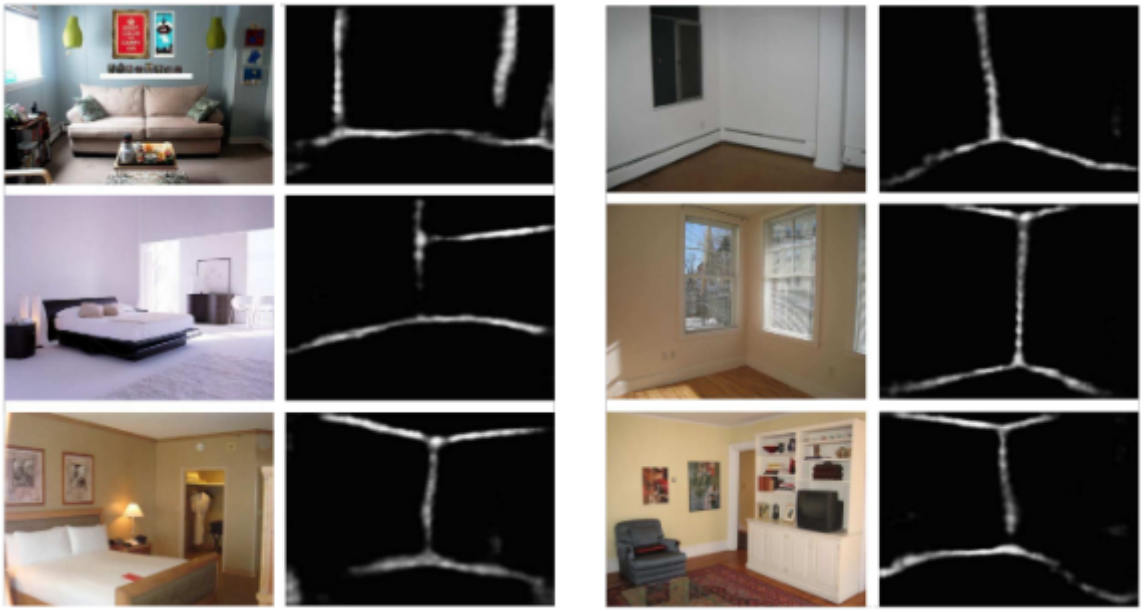


FIGURE 1.52 – Un exemple d'extraction de structure réalisé par réseaux de neurones [Mallya and Lazebnik, 2015].

PanoRoom [Fernandez-Labrador et al., 2018] utilise une architecture d'encodeur-décodeur avec une base convolutive ResNet-50. Des exemples de prédictions sont visibles sur la FIGURE 1.53. LayoutNet [Zou et al., 2018] utilise également une architecture d'encodeur-décodeur. La FIGURE 1.54 illustre des exemples de prédiction réalisées grâce à LayoutNet.



FIGURE 1.53 – Illustration de prédiction réalisée par le réseau PanoRoom [Fernandez-Labrador et al., 2018]. On retrouve en vert la vérité terrain et en jaune les prédictions produites par le réseau.



FIGURE 1.54 – Illustration de prédiction réalisée par le réseau LayoutNet [Zou et al., 2018]. On retrouve en vert la vérité terrain et en orange les prédictions réalisées par le réseau.

Systèmes embarqués

Un dernier aspect qui nous préoccupe, est la capacité de ces modèles prédictifs à être embarqués dans des dispositifs mobiles. Nous avons besoin pour cela que les réseaux de neurones soient capables de réaliser les prédictions en temps réel et en consommant peu de ressources.

Il existe à ce jour plusieurs réseaux de neurones dont l'architecture est plus légère. On peut notamment citer le réseau [Howard et al., 2017]. Ce réseau fait appel à des "*depthwise separable convolution*" qui permet de gérer moins de paramètres et donc moins de calculs que les convolutions classiques. MobileNet est un type de réseau conçu pour être embarqué sur téléphone, ce qui en fait un très bon candidat. Il y a également l'architecture Sonnet [12] qui utilise "`tensorflow.js`" afin de pouvoir réaliser les prédictions dans une interface web et donc nécessite le moins de calculs possibles afin de fonctionner en temps réel. Enfin, la famille compacte de réseaux EfficientNet [Tan and Le, 2019], dont l'architecture a été optimisée par renforcement, permet également de faire des prédictions en temps réel.

Les méthodes de prédiction présentées dans la section 1.5.2 possèdent toutes des bases convolutives. Il est donc possible de modifier ces bases par les réseaux de neurones nommés ci-dessus (MobileNet, EfficientNet) ce qui en ferait des réseaux bien plus légers.

En plus de la réduction de la taille des réseaux (et donc de la diminution du nombre de calculs), il est possible de réduire encore plus les coûts de prédictions grâce au mécanisme de "*pruning*" et de conversion des nombres manipulés en float16 (l'espace mémoire occupé par chaque nombre dans le réseau est divisé par quatre). Ces techniques peuvent être mises en place avec la librairie TFLite [13].

12. `https://github.com/deepmind/sonnet`
13. `https://www.tensorflow.org/lite/guide?hl=fr`

En résumé

Nous avons présenté dans ce chapitre les éléments principaux autours desquels vont s'articuler les travaux qui sont présentés dans ce manuscrit. Nous nous sommes dans un premier temps intéressés aux implants réels, avec leurs caractéristiques (avantages/défauts). La difficulté d'accès aux patients implantés nous a amenés à discuter de l'utilisation d'un simulateur de vision prothétique. Nous avons présenté les problématiques soulevées par l'utilisation de simulateur de vision prothétique, et avons dressé une liste (non-exhaustive) de simulateurs utilisés dans la littérature depuis 1992 jusqu'en 2022. Pour conclure cette partie, nous avons brièvement introduit les réseaux de neurones. Les récentes avancées des capacités de prédictions et de compacités de ces réseaux permettent d'intégrer une analyse d'image avancée sur le flux vidéo issue de la caméra placée sur le non-voyant, et donc de plus facilement extraire les informations jugées importantes pour une bonne restitution phosphénique.

Étant donné que nous avons besoin de pouvoir simuler une restitution phosphénique, nous avons dû développer un simulateur de vision prothétique. Nous avons besoin de pouvoir simuler aussi bien des environnements intérieurs qu'extérieurs, et de permettre aux sujets de pouvoir se déplacer. Notre choix s'est porté sur le moteur de jeu Unity 3D qui permet de représenter des environnements de complexité variable et d'intégrer un module de vision phosphénique. Le chapitre 2 présentera Unity 3D, puis la méthodologie de construction de la vision phosphénique, les méthodes d'extraction de caractéristiques des images issues de la scènes et enfin le fonctionnement du simulateur.

Chapitre 2

Simulateur de vision prothétique

Sommaire

2.1 Unity 3D 54
2.1.1 Présentation générale 54
2.1.2 Aspect technique 55
2.2 Construction de la représentation phosphénique 58
2.2.1 Résolution de l'implant 59
2.2.2 Taux d'électrodes défectueuses 59
2.2.3 Géométrie variable des phosphènes 62
2.2.4 Espacement inter-phosphènes 62
2.2.5 Nombre de niveaux de gris variable 63
2.2.6 Flou gaussien 64
2.2.7 Tous les défauts en même temps 64
2.3 Extraction des caractéristiques et rendu 64
2.3.1 "Scoreboard" 66
2.3.2 Contour brut 68
2.3.3 Segmentation sémantique et extraction d'objets 68
2.3.4 Contour amélioré 69
2.3.5 Structure seule 70
2.3.6 Méthode combinée 71
2.4 Fonctionnement du simulateur 72
2.4.1 Système de caméra 72
2.4.2 Sélection des paramètres 74
2.4.3 Comparaison des méthodes de restitutions 74

Dans ce chapitre, nous présentons le fonctionnement et les caractéristiques principales de notre simulateur de vision prothétique. À l'origine, il était prévu de développer tout le système en réalité virtuelle et de réaliser les expériences dans une salle dédiée : pendant les six premiers mois de la thèse, le dispositif était en place et quelques expériences ont pu être menées avant la crise sanitaire. Pour cela, nous avions fait l'acquisition d'un casque HTC Vive Pro Eye (meilleur casque disponible sur le marché à l'époque). Pour s'adapter à l'évolution des conditions de travail, nous avons dû transformer le simulateur sous forme d'un exécutable. Cette version a servi à conduire la première expérimentation, mais présentait des soucis de compatibilité. Les problèmes issus de la première étude nous ont amenés à proposer une seconde version directement en ligne. Le déploiement en ligne présente l'avantage de faciliter l'accès aux participants : il leur suffit de cliquer sur un lien pour participer à l'étude.

Nous présentons dans un premier temps le moteur de jeu vidéo Unity 3D (2.1). Ensuite, nous expliquons comment nous construisons une représentation phosphénique à partir d'une image donnée (2.2) et les techniques d'extraction de caractéristiques utilisées dans le simulateur (2.3). Finalement, nous montrerons l'interface d'utilisation du simulateur (2.4).

2.1 Unity 3D

2.1.1 Présentation générale

Unity est un moteur de jeu multi-plateforme (smartphone, ordinateur, consoles de jeux vidéo et Web) développé par Unity Technologies. Il est l'un des plus répandus dans l'industrie du jeu vidéo, aussi bien pour les grands studios que pour les indépendants du fait de sa rapidité à construire des prototypes de jeux/simulations, et qu'il permet de sortir les jeux sur tous les supports (consoles, ordinateur, web, smartphone). L'éditeur est construit sur une base de code en C++ et C# et la programmation se fait principalement en C#.

Il a notamment été utilisé dans la création de Pokémon GO, Ori and the Blind Forest, Furi, Haven ou encore Hearthstone : Heroes of Warcraft qui sont des références connues dans le milieu du jeu vidéo. À titre d'exemple, Pokémon Go a rapporté 642 millions de dollars en 2021. Les principales qualités de Unity 3D sont la présence d'un moteur de simulation graphique sophistiqué ainsi que la présence d'un moteur de simulation physique robuste.

Les Figures 2.1 et 2.2 illustrent illustrent respectivement le jeu Furi et le jeu Haven, tout deux développés par le studio The Game Bakers [1].

1. https://www.thegamebakers.com/



FIGURE 2.1 – Illustration du jeu Furi, développé par The Game Bakers. Le développement a été fait sous Unity 3D.



FIGURE 2.2 – Illustration du jeu Haven, développé par The Game Bakers. Le développement a été fait sous Unity 3D.

2.1.2 Aspect technique

Moteur graphique

Les fonctionnalités graphiques d'Unity permettent de contrôler l'apparence de l'application et sont hautement personnalisables. Il est possible d'utiliser les fonctionnalités graphiques d'Unity pour créer des graphismes très réalistes et optimisés. Dans notre cas, la quantité de ressources gratuites sur Internet nous a permis de construire des scènes adaptées à nos expériences composées d'une variété d'objets que ce soit en intérieur ou en extérieur.

Parmi les éléments graphiques sur lesquels il est possible d'agir, on retrouve le "*pipeline*" de rendu, les caméras, les effets de post-traitement (comme par exemple l'*anti-aliasing*,

le *bloom* ou le brouillard), les jeux de lumières, les textures et matériaux, les systèmes de particules et les espaces colorimétriques.

Les FIGURES 2.3 , 2.4 [2] et 2.5 illustrent respectivement les capacités de Unity 3D en termes de gestion de lumières et de modélisation d'environnements intérieurs et extérieurs.



FIGURE 2.3 – Capture d'écran du jeu "The Heretic", réalisé sous Unity 3D.



FIGURE 2.4 – Environnement 3D d'intérieur modélisé entièrement en Unity 3D.

Moteur physique

Unity 3D permet également de simuler la physique dans l'environnement afin de garantir que les objets accélèrent correctement et répondent aux collisions, à la gravité et à diverses autres forces. Par ce biais, il est possible de manipuler avec précision les interactions entre les différents éléments de la simulation.

2. https://assetstore.unity.com/packages/3d/environments/urban/archvizpro-interior-vol-6-120489



FIGURE 2.5 – Exemple de scène extérieure modélisée sous Unity 3D. Extrait du jeu "Escape from Tarkov".

On peut notamment citer le jeu "Kerbal Space Program" qui est un simulateur de vol spatial permettant au joueur de construire sa propre fusée, son propre avion, son propre satellite artificiel, ou même sa propre station spatiale puis de les lancer dans l'espace. Certaines institutions et entreprises du secteur spatial ont manifesté de l'intérêt pour Kerbal Space Program. La FIGURE 2.6 illustre le mode de création de fusée issu du jeu Kerbal Space Program. C'est le cas de la NASA qui a participé à la mise à jour "Asteroid Redirect Mission", largement inspirée de son propre projet de capture d'astéroïde. En 2020, l'équipe de développement lance un partenariat avec l'Agence Spatiale Européenne pour le développement de la mise à jour "Shared Horizons", qui permet une reproduction du lanceur européen Ariane 5 ainsi que des missions d'exploration Rosetta[3].

Scripts et ECS

Les scripts sont un élément essentiel dans l'utilisation de Unity. Ils sont représentés par du code C# sous forme de méthodes. La plupart des applications ont besoin de scripts pour répondre aux entrées du joueur et pour faire en sorte que les événements liés à la simulation se produisent au moment voulu. En outre, les scripts peuvent être utilisés pour créer des effets graphiques, contrôler le comportement physique des objets ou même mettre

3. https://fr.wikipedia.org/wiki/Kerbal_Space_Program#cite_note-87



FIGURE 2.6 – Image tirée du jeu Kerbal Space Program. On se trouve ici dans le mode de création d'une fusée. Les items sur la gauche sont les composants que l'on peut attacher à la fusée.

en œuvre un système d'IA autonome.

Unity 3D utilise un *Entity Component System* pour la gestion de ses objets. Un ECS privilégie la composition de comportements par rapport à l'héritage de classes, ce qui signifie que chaque entité est définie non pas par un type, mais par les composants qui lui sont associés. La conception de la manière dont les composants sont associés aux entités dépend du système de composants d'entités utilisé.

Dans la prochaine section (2.2), nous allons détailler les processus mis en place dans notre simulateur, développé en utilisant Unity 3D, qui permettent de construire une représentation phosphénique, à partir d'une image issue d'une caméra placée dans la scène virtuelle. C'est cette méthodologie qui sera appliquée aux extractions de caractéristiques de la section 2.3.

2.2 Construction de la représentation phosphénique

Pour obtenir une restitution visuelle, il faut décomposer le processus en deux étapes. On doit dans un premier temps être capable de construire une image contenant les dif-

férents éléments à mettre en évidence. Dans le cas d'une segmentation sémantique (mise en évidence des objets), il s'agit de changer la couleur de chaque pixel, qui correspond à un objet, en fonction de sa classe. Ainsi, si le pixel est associé à une voiture, il sera d'une certaine couleur (par exemple bleue), et si le pixel ne correspond à aucun objet, il sera noir. Une fois cette image d'éléments d'intérêt créée, il faut lui appliquer un masque phosphénique, qui va permettre de la transformer en image phosphénique, c'est-à-dire sous forme de phosphènes.

Notre simulateur permet de gérer plusieurs paramètres :

— la résolution de l'implant (c'est-à-dire le nombre d'électrodes simulées) (2.2.1) ;
— le taux d'électrodes défectueuses (2.2.2) ;
— la géométrie des phosphènes (2.2.3) ;
— l'espacement inter-phosphènes (2.2.4) ;
— un nombre de niveaux de gris limité (2.2.5) ;
— la présence d'un flou gaussien (2.2.6) ;

Nous avons choisi de ne pas traiter certains phénomènes tels que l'évanescence ou la persistance phosphénique, car ces éléments n'interviendront pas dans les expérimentations conduites. En revanche, il est tout à fait possible d'intégrer ces fonctionnalités dans le simulateur de par sa modularité.

Dans cette section, nous présentons la méthode de construction des masques phosphéniques, et dans la section suivante (2.3), nous présenterons différentes techniques d'extraction d'éléments caractéristiques.

2.2.1 Résolution de l'implant

Le premier paramètre réglable est la résolution de l'implant. Cela correspond au nombre de phosphènes présents sur l'implant simulé. L'organisation des phosphènes suit une grille rectangulaire régulière. L'image issue de la caméra est découpée et chaque zone ainsi créée va ensuite accueillir un phosphène. Sur la FIGURE 2.7, on peut voir deux environnements à différentes résolutions d'implant. La résolution est modifiable directement dans le simulateur.

2.2.2 Taux d'électrodes défectueuses

Lors de l'implantation, des matrices d'électrodes, entre 10% et 30% des électrodes se brisent ou sont défectueuses. On modélise ce phénomène en désactivant une proportion des électrodes. Le taux d'électrodes défectueuses est un paramètre réglable depuis le simulateur. La FIGURE 2.8 illustre plusieurs taux d'électrodes défectueuses.

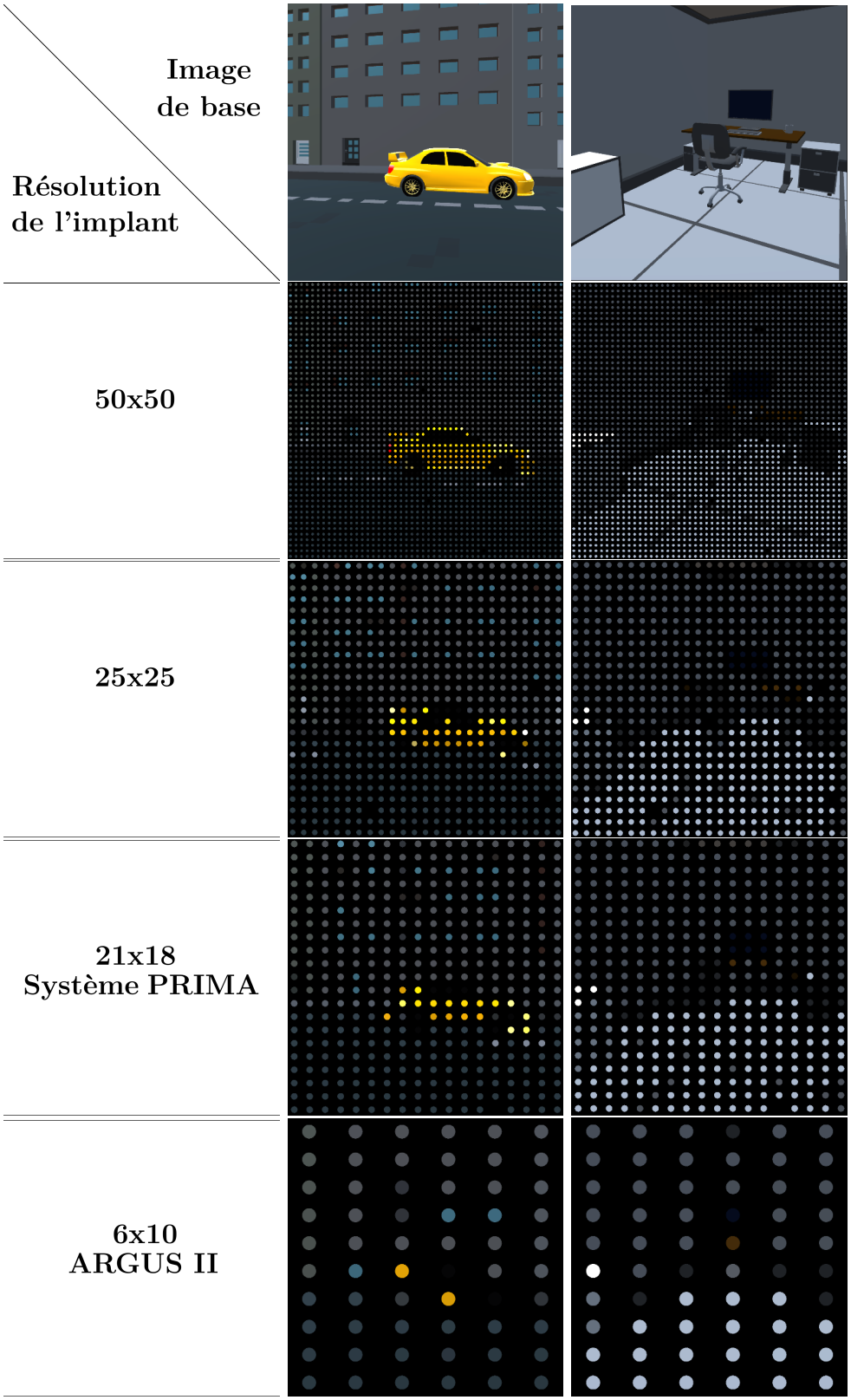


FIGURE 2.7 – Exemple d'implant simulé de différentes résolutions. Deux environnements (extérieur et intérieur) sont présentés. Les résolutions montrées sont 50x50, 25x25, 21x18 (résolution du système PRIMA, 6x10 (résolution de l'ARGUS II).

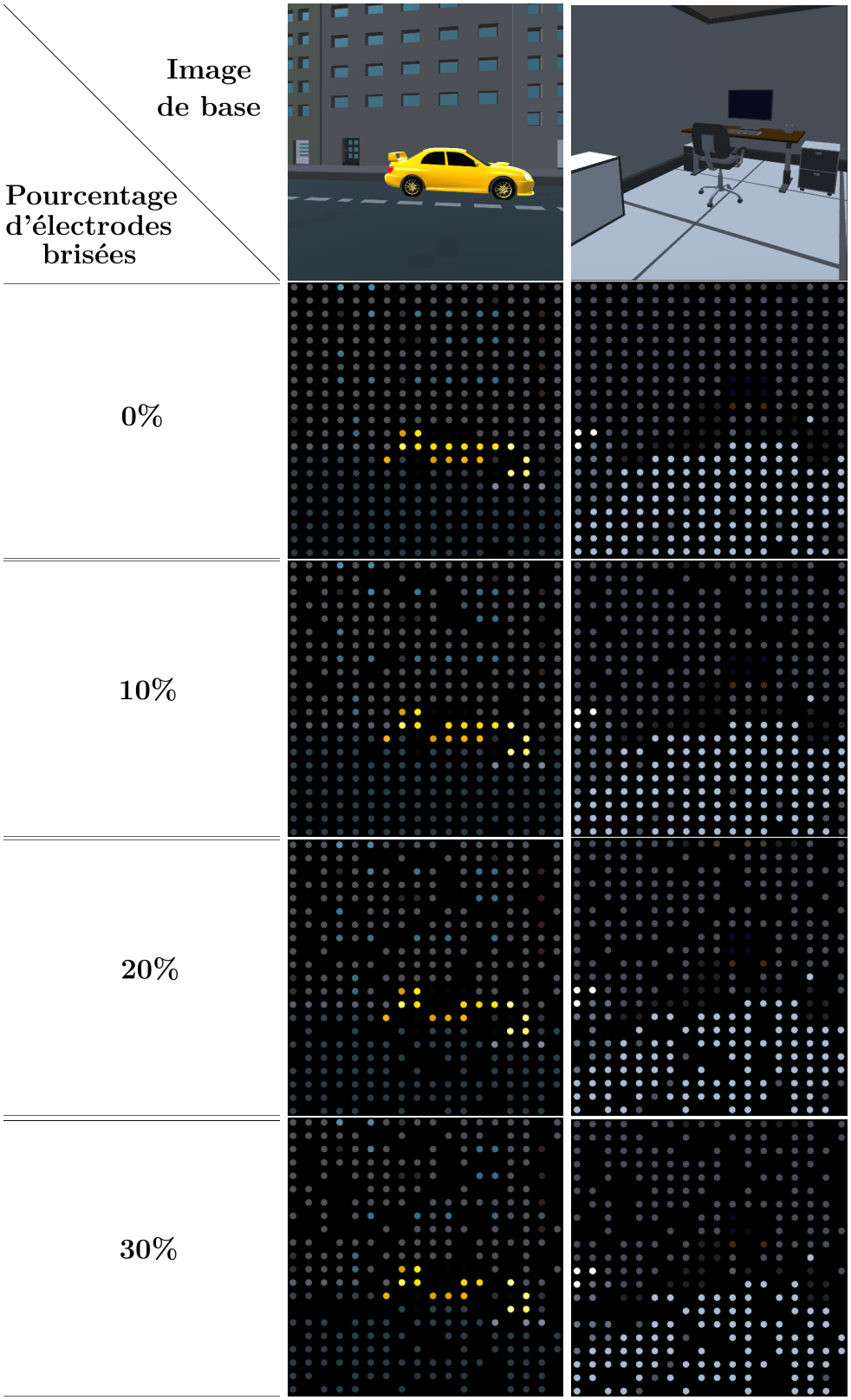


FIGURE 2.8 – Exemple d'implant simulé avec un taux d'électrodes défectueuses variable. Deux environnements (extérieur et intérieur) sont présentés. Les taux montrés sont 0%, 10%, 20% et 30%. La résolution utilisée est de 21x18 (Système PRIMA).

2.2.3 Géométrie variable des phosphènes

Nous avons choisi de modéliser les phosphènes par une représentation circulaire. On peut envisager de représenter les phosphènes sous une forme arbitrairement complexe à condition qu'il soit possible de décrire cette géométrie de manière implicite (par exemple sous forme hexagonale). Nous avons ajouté la possibilité de faire varier le rayon des cercles. Les rayons peuvent varier entre 0,15 et 0,25 fois la plus petite dimension d'une cellule phosphénique (emplacement prévu pour recevoir un phosphène). La FIGURE 2.9 illustre des exemples de variabilité des rayons.

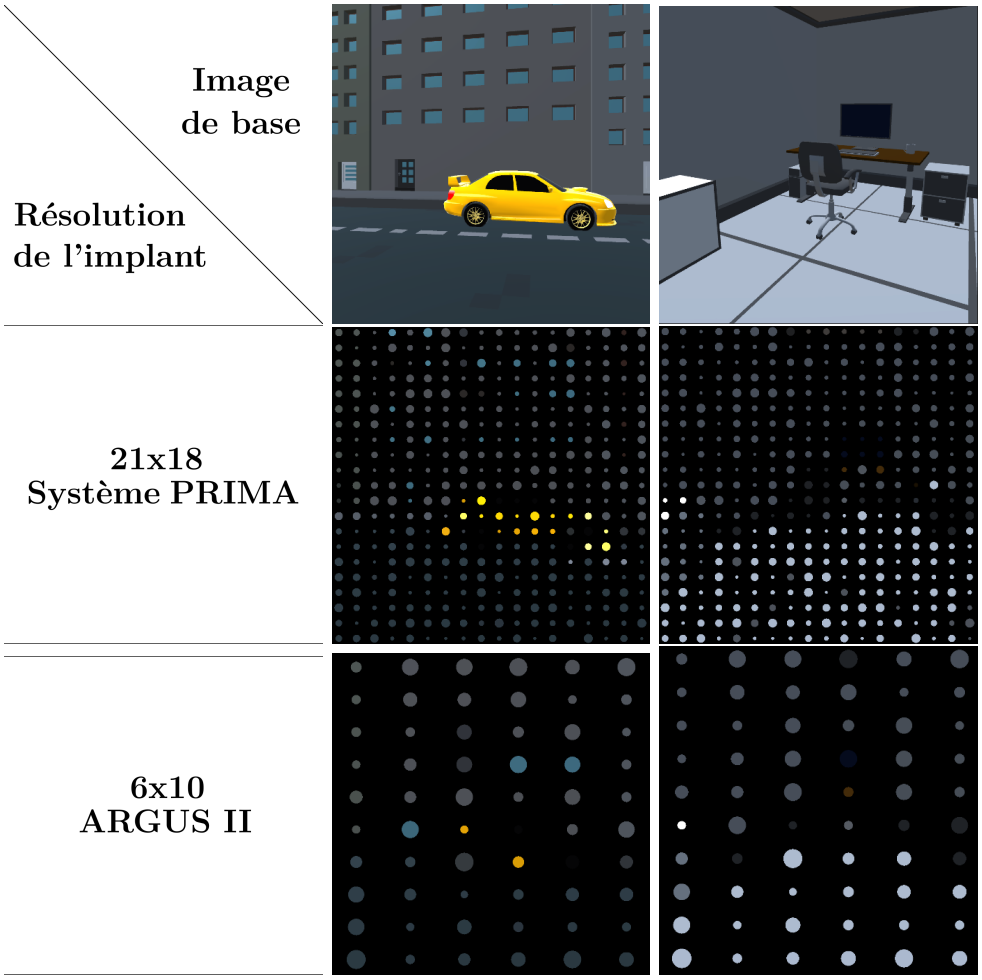


FIGURE 2.9 – Exemple d'implants simulés de résolution variable. Deux environnements (extérieur et intérieur) sont présentés. Cette figure met en évidence la possibilité de définir des rayons de phosphènes variables au sein d'un implant.

2.2.4 Espacement inter-phosphènes

Notre simulateur permet également d'introduire des déformations dans la cartographie rétinotopique. Cela est fait dans des proportions modérées, en décalant les phosphènes

générés à l'intérieur de la cellule phosphénique qui leur a été attribuée. Il suffit pour cela de déplacer le centre du phosphène verticalement et horizontalement. Chaque phosphène subit ainsi un déplacement différent. Des exemples d'espacement inter-phosphènes sont présentés sur la FIGURE 2.10.

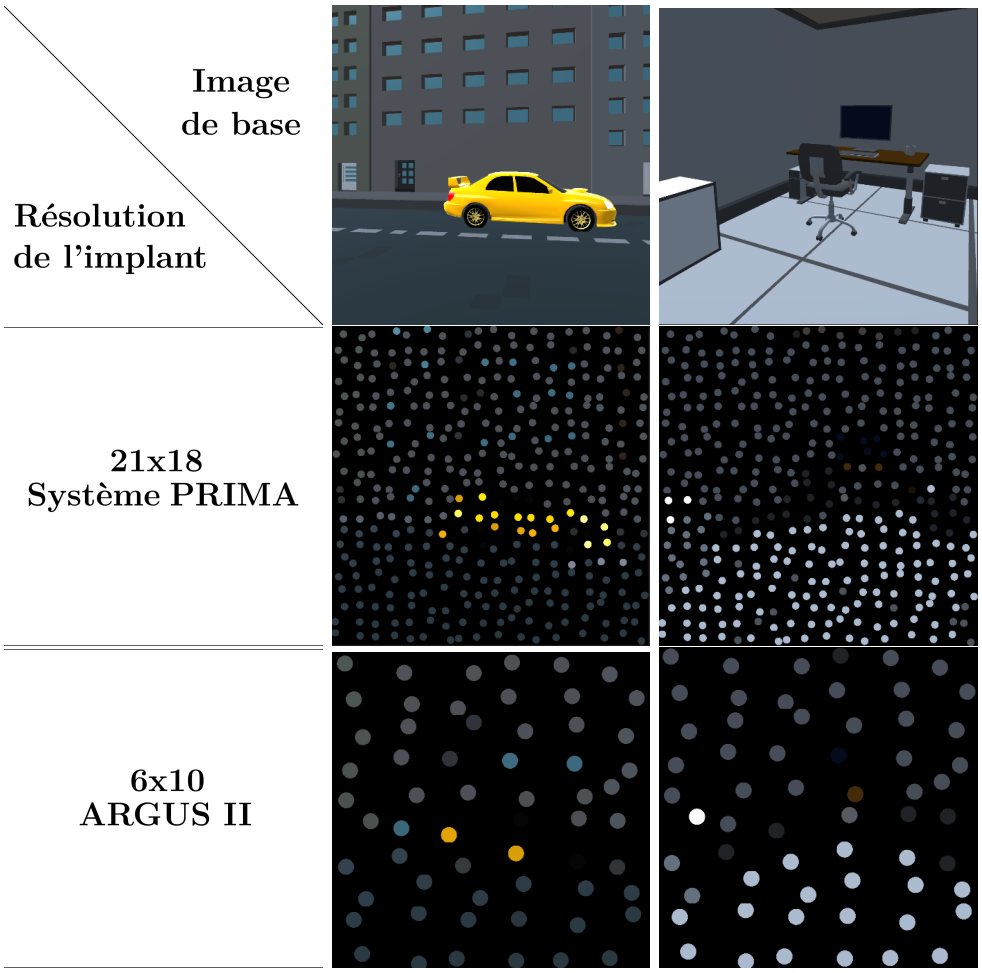


FIGURE 2.10 – Exemple d'implants simulés de résolution variable. Deux environnements (extérieur et intérieur) sont présentés. Cette figure met en évidence la possibilité de définir des rayons de phosphènes variables au sein d'un implant.

2.2.5 Nombre de niveaux de gris variable

Comme présenté plus haut dans l'état de l'art (chapitre 1), les couleurs ne sont pas toujours perçues par les sujets implantés. La majorité des sujets perçoivent des phosphènes d'intensité variable, principalement de couleur blanche. Selon le type d'implant utilisé, le nombre de niveaux de gris différents perceptible varie entre 4 et 8. Ajoutons à cela que les niveaux de gris varient par électrodes, et ne sont pas uniformément répartis entre 0 et 255. Bien que nous ayons choisi de mettre en place la variabilité du nombre de niveaux de gris

pour une même matrice, nous avons décidé de ne pas faire varier l'étalement de ces valeurs entre 0 et 255 autrement qu'uniformément. Ce choix est motivé par le fait que nous avons besoin de garder un certain contrôle sur la réalisation des tâches des expérimentations des chapitres 3 et 4. Pour observer des résultats significatifs, il faut que la tâche ne soit ni trop complexe, ni trop difficile. La FIGURE 2.11 illustre cette fonctionnalité pour plusieurs nombres de niveaux de gris.

2.2.6 Flou gaussien

Enfin, notre simulateur permet d'activer un flou gaussien global sur l'image pour effacer les contours nets et peu réalistes. Pour ce faire, on applique un filtre gaussien sur toute l'image, horizontalement et verticalement. La FIGURE 2.12 montre l'application du flou gaussien sur un implant simulé de taille 21x18.

2.2.7 Tous les défauts en même temps

Finalement lorsque tous les problèmes présentés ci-dessus sont activés en même temps, on observe des restitutions phosphéniques comme celles montrées sur la FIGURE 2.13. On constate que la compréhension de la scène est plus complexe que dans le cas où les défauts ne sont pas tous présents en même temps, particulièrement dans le cas de l'implant simulé ARGUS II.

Nous venons de présenter comment les restitutions phosphéniques sont modélisées dans notre simulateur. La sous-section suivante (2.3) se concentre sur l'extraction de caractéristiques depuis l'image d'entrée acquise par la caméra dans la simulation. Nous détaillerons ainsi les méthodes classiquement utilisées à savoir *scoreboard* et *contour brut*, puis nous présenterons des techniques plus sophistiquées, comme par exemple l'extraction d'objets, pouvant faire appel à des réseaux de neurones (présentés en 1.5).

2.3 Extraction des caractéristiques et rendu

Nous détaillons ici les différentes méthodes d'extraction de caractéristiques et de rendu utilisés tout au long de ce manuscrit. Dans un premier temps, nous détaillons deux techniques classiques, à savoir, la méthode *Scoreboard* (2.3.1) et la méthode *contour brut* (2.3.2). Ensuite, nous présentons quatre techniques plus sophistiquées : *extraction d'objets* (2.3.3), *extraction de contours* (2.3.4), extraction de structure (2.3.5) et enfin *extraction combinée* (2.3.6).

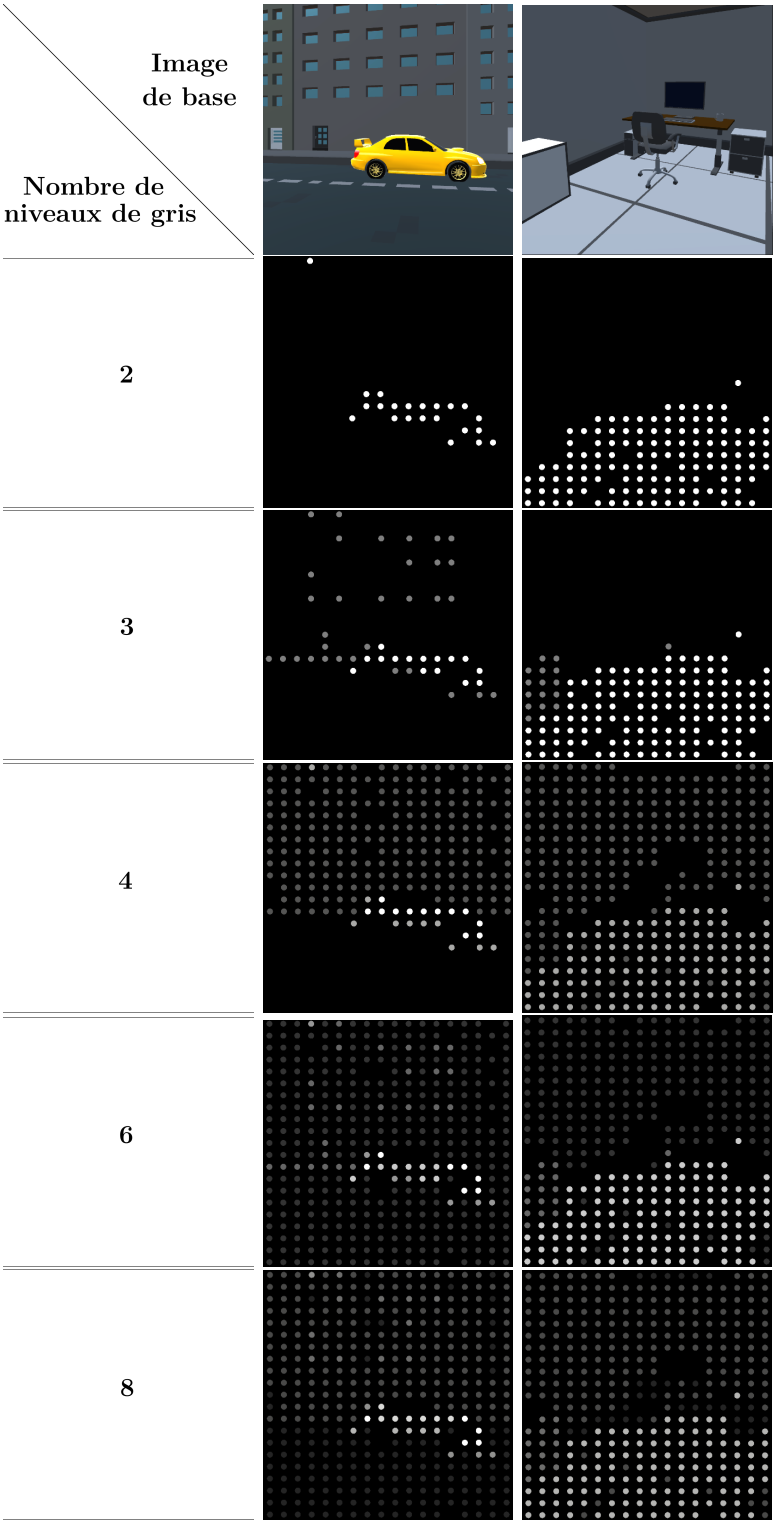


FIGURE 2.11 – Exemple d'implants simulés avec plusieurs niveaux de gris (2, 3, 4, 6, 8). Deux environnements (extérieur et intérieur) sont présentés. L'implant simulé est de résolution 21x18 (Système PRIMA)

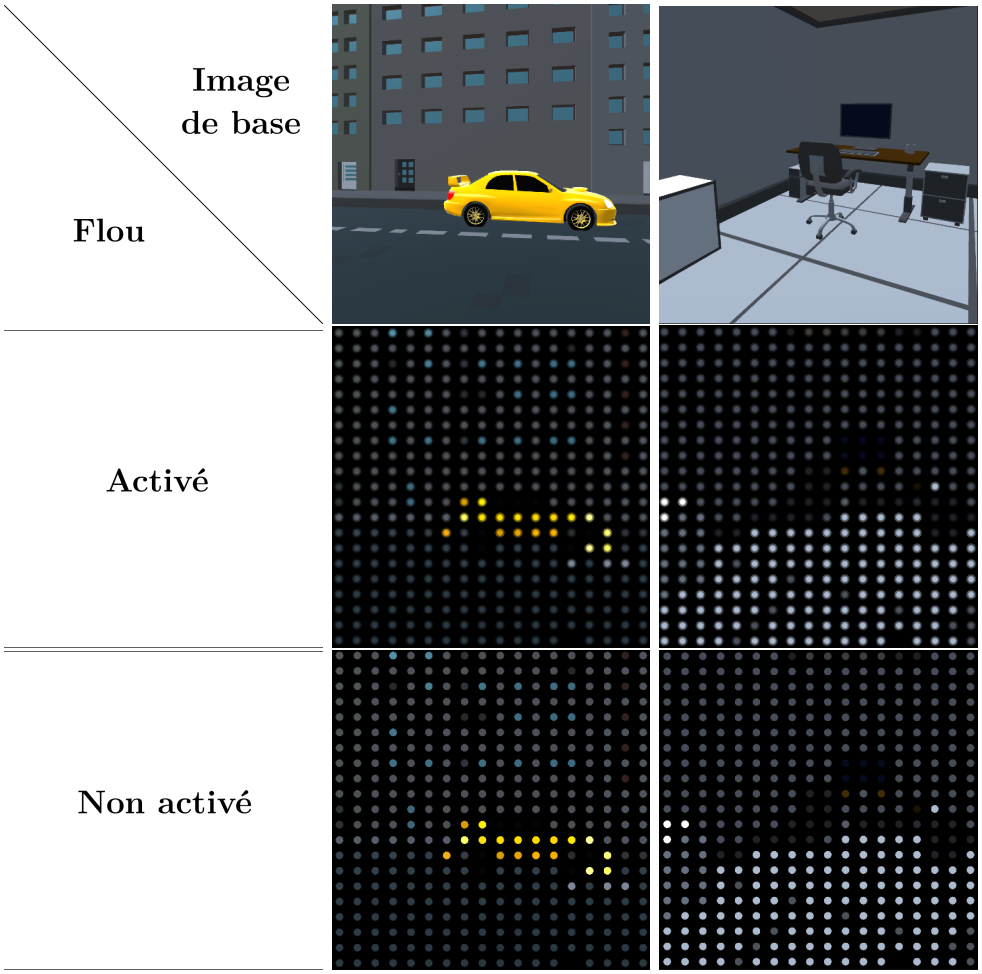


FIGURE 2.12 – Exemple d'implants simulés avec et sans activation du flou gaussien. Deux environnements (extérieur et intérieur) sont présentés. L'implant simulé est de résolution 21x18 (Système PRIMA).

2.3.1 "Scoreboard"

La méthode *Scoreboard* est la méthode historique de rendu phosphénique. Elle consiste en la récupération des niveaux de gris moyens dans chacune des zones recevant un phosphène et l'attribution de cette intensité lumineuse au phosphène correspondant.

Dans un premier temps, on cherche à récupérer la couleur moyenne dans chaque cellule contenant le phosphène à afficher sous forme d'image. Le mécanisme de récupération est illustré sur la FIGURE 2.14. Ensuite, on transforme cette image en niveau de gris. Après application du sélecteur du nombre d'intensités désiré, on réalise la transformation phosphénique.

Cette méthode est la plus simple que nous utiliserons étant donné qu'elle ne nécessite pas de traitement d'images avancé. En revanche, elle comporte des défauts importants : difficulté de comprendre le lien entre intensité des phosphènes et éléments représentés,

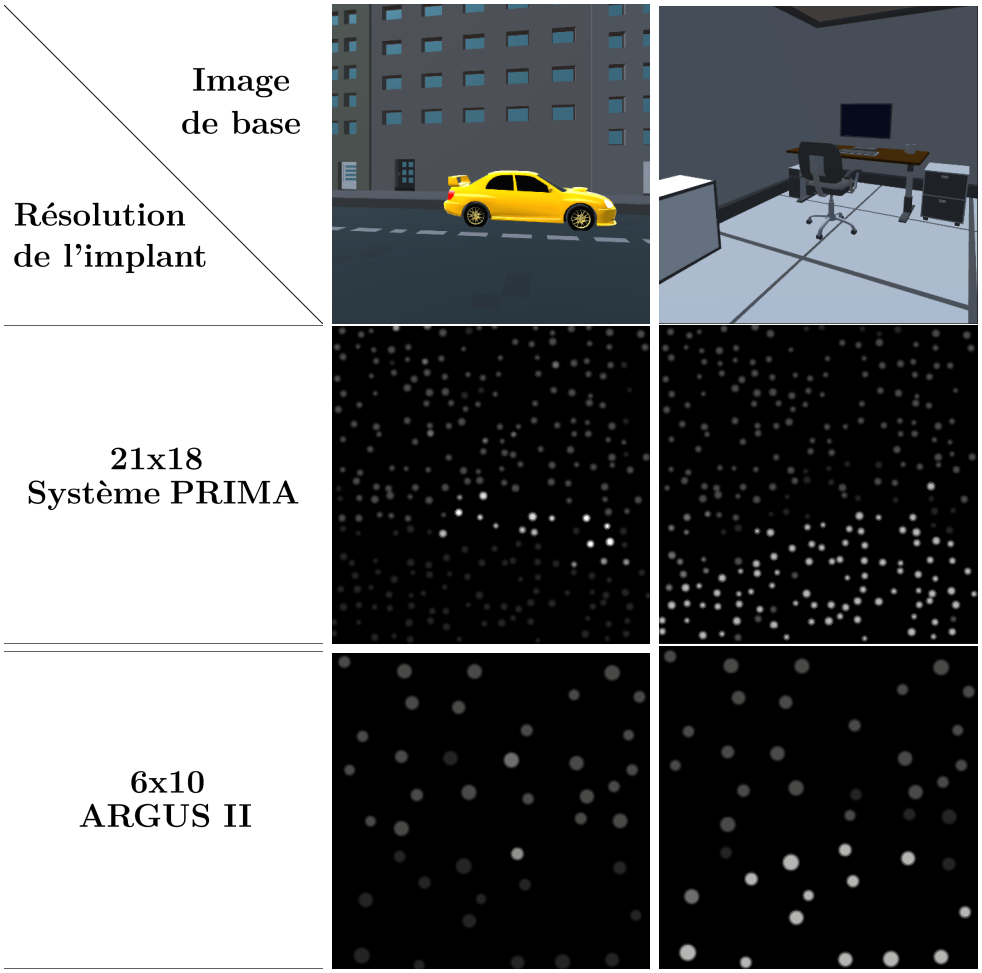


FIGURE 2.13 – Exemple d'implants simulés de résolution 21x18 et 6x10. Deux environnements (extérieur et intérieur) sont présentés. Tous les défauts sont modélisés. Ici, on a 30% de *dropout* et 8 niveaux de gris.

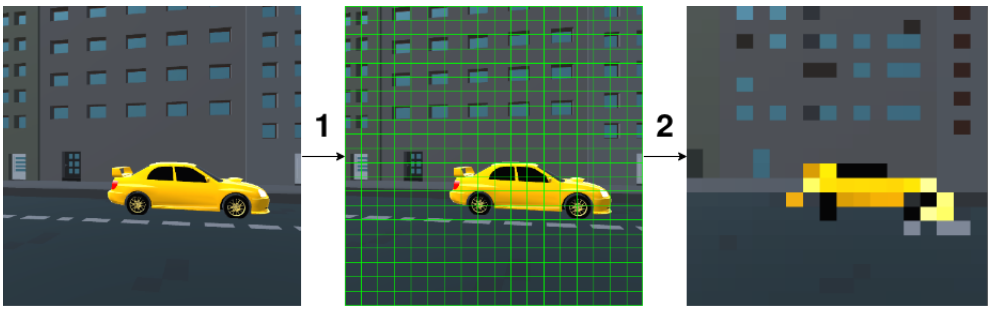


FIGURE 2.14 – Illustration de la méthodologie *scoreboard*. Les étapes de construction sont les suivantes : (1) - Partitionnement de l'image en carrés de même taille (étape de mise à l'échelle de l'implant), (2) - Pour chaque pixel, on associe l'intensité moyenne de la cellule à laquelle le pixel appartient.

forte sensibilité à l'éclairage de la scène.

2.3.2 Contour brut

Cette méthode est basée sur une détection des contours grâce à un filtre de Sobel ou de Canny. Dans un premier temps, on calcule une carte de Sobel à partir de l'image de base. Le processus consiste à réaliser des opérations de convolutions avec des filtres particuliers. Ensuite, on diminue la résolution en réalisant des moyennages successifs pour obtenir une image décrivant les contours en plus faible résolution.

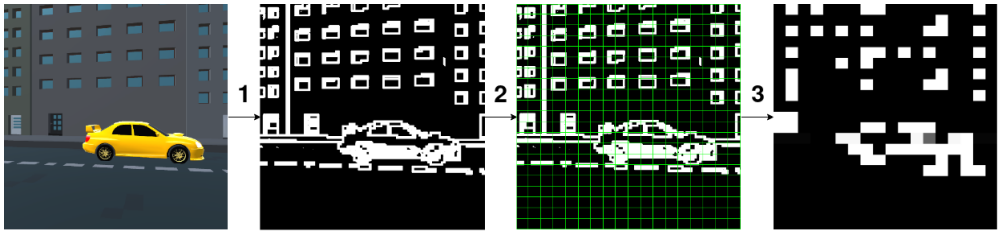


FIGURE 2.15 – Illustration de la méthode *contour brut*. Les étapes de construction sont les suivantes : (1) – Extraction des contours via l'application d'un filtre de Canny glissant sur l'image de base (ici, une dilatation des contours a également été appliquée pour bien mettre en évidence ces derniers), (2) - Découpe des contours en cellules de même taille (étape de mise à l'échelle de l'implant), (3) - Pour chaque cellule, on remplace l'intensité des pixels qui le composent par 255 si plus de la moitié des pixels sont blancs dans cette cellule et par 0 sinon.

Comme on peut le voir sur la FIGURE 2.15, cette méthode permet de mettre en évidence certains éléments structurels. On peut notamment distinguer la silhouette de la voiture sur la quatrième image. La méthode de détection de contours avec filtre de Canny repose sur le choix de plusieurs hyper-paramètres liés au filtre (seuil de détection - ici la moitié des pixels d'une cellule, taille du filtre - dans notre cas 3x3). Un filtre avec certains paramètres peut donc fonctionner très bien dans un contexte donné et ne plus extraire de structure dans un autre. Ce n'est pas une technique très robuste dans le cas général.

Dans les méthodes suivantes, nous n'expliciterons plus la phase de découpe en cellules.

2.3.3 Segmentation sémantique et extraction d'objets

Afin de pouvoir extraire les différents objets de l'image issue de la caméra, il faut utiliser un algorithme de segmentation. Étant donné que nous sommes dans un contexte de simulation, on connaît déjà la classe de chaque objet, ce qui nous donne de manière immédiate la segmentation. En pratique, nous étiquetons tous les éléments qui peuvent apparaître dans notre scène. Cela inclut aussi bien les éléments de structure (comme les murs, le sol, le plafond) que les objets (voiture, porte, chaise, table, …). Une fois les étiquettes asso-

ciées aux objets, on est en mesure de récupérer en temps réel la carte de segmentation sémantique associée à l'image issue de la caméra. Cette carte dénote pour chaque pixel son appartenance à une classe d'objets en utilisant la couleur comme mode de représentation. Un exemple d'annotation est montré sur la FIGURE 2.16.



FIGURE 2.16 – Illustration de la méthode de segmentation sémantique. Les étapes de construction sont les suivantes : (1) – Calcul de la carte de segmentation, soit grâce à des réseaux de neurones, soit en utilisant le mécanisme d'étiquettes de notre simulateur, (2) - Mise à l'échelle de l'implant simulé (méthode analogue à scoreboard, 2.3.1)

Cette technique permet de mettre en évidence les différents éléments qui constituent l'image. Dans le cas de notre simulateur, les objets sont étiquetés automatiquement dès qu'ils sont instanciés dans le monde. Pour obtenir cette carte de segmentation sémantique, il suffit de changer la texture des objets directement dans la simulation et aucun traitement d'image n'est à faire. Dans un cas réel d'utilisation, il faudrait utiliser un réseau de neurones, tel que présenté dans le premier chapitre 1.5, qui soit capable de prédire ces cartes de segmentation sémantique en temps réel.

Une variante de ce rendu consiste à ne faire apparaître que les éléments qualifiés d' "objets", c'est-à-dire tout sauf les bâtiments, murs, sol, plafond, trottoirs. Dans ce cas, il faut au préalable filtrer les classes de segmentation sémantique. Cette méthode est proche de la distinction *thing/stuff* de MSCOCO (qui permet de distinguer les éléments importants d'une image). La FIGURE 2.17 illustre cette méthode qui est très efficace pour mettre en évidence les objets qui constituent la scène.

On observe dans ce cas que la porte de gauche semble disparaître. Ce comportement est dû au mécanisme de mise à la résolution de l'implant, qui en faisant les moyennages locaux, fait disparaître cette zone. La porte est située à cheval sur plusieurs cellules, et lors du moyennage des couleurs de chaque cellule, la proportion de pixels appartenant à la porte dans chacune des cellules qu'elle couvre est inférieure à la proportion de pixels correspondant aux murs.

2.3.4 Contour amélioré

Ce rendu a pour objectif de mettre en évidence les jointures entre les différents éléments qui composent la scène. C'est une amélioration d'un rendu précédent (*contour brut*). Pour

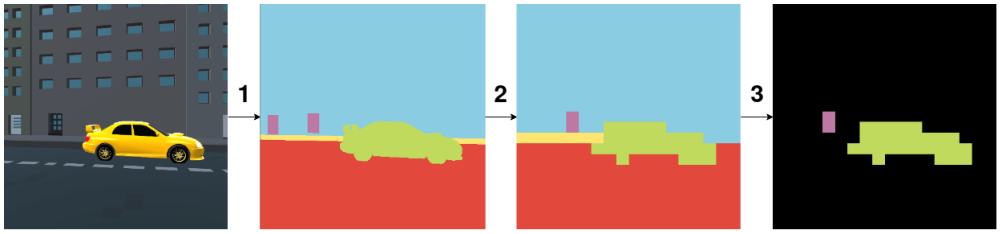


FIGURE 2.17 – Illustration de la méthode de *segmentation sémantique d'objets*. Les étapes de construction sont les suivantes : (1) – Calcul de la carte de segmentation, soit grâce à des réseaux de neurones, soit en utilisant le mécanisme d'étiquettes de notre simulateur, (2) – Mise à l'échelle de l'implant simulé (méthode analogue à scoreboard, 2.3.1), (3) – Filtrage des classes d'objets (dans notre cas : voiture et portes).

cela, on se base sur la carte de segmentation sémantique, puis on calcule des cartes de gradients horizontaux et verticaux que l'on combine en une carte de contours. La FIGURE 2.18 illustre ce procédé.

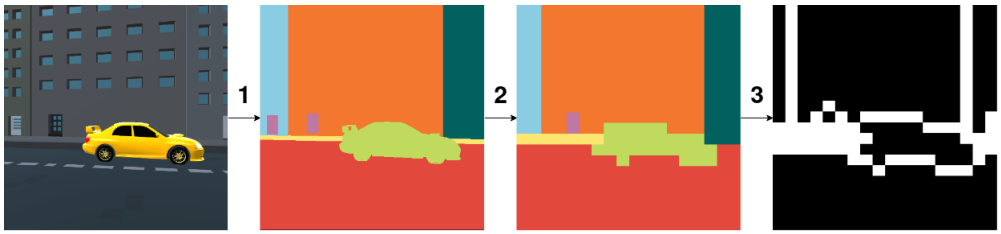


FIGURE 2.18 – Illustration de la méthode d'*extraction de contours améliorée*. Les étapes de construction sont les suivantes : (1) – Calcul de la carte de segmentation, soit grâce à des réseaux de neurones, soit en utilisant le mécanisme d'étiquettes de notre simulateur. Ici, les classes de segmentation sont propres à chaque entité de la scène. (2) – Mise à l'échelle de l'implant simulé (méthode analogue à *Scoreboard*, 2.3.1), (3) – Application des gradients (horizontal et vertical) pour détecter les changements d'objets.

Dans cette technique, on a besoin que chaque entité possède une classe distincte, on peut alors dans ce cas mettre en évidence toutes les jointures entre deux éléments. Cette technique est une amélioration de la méthode *contour brut*, car elle ne nécessite pas de gestion d'hyper-paramètres comme la taille d'un filtre par exemple. Comme pour la méthode précédente (*extraction d'objets*), cette technique repose sur la prédiction d'une carte de segmentation sémantique, qui peut être réalisée par un réseau de neurones. On note que la porte de gauche est aussi perdue dans le processus de manière analogue à 2.3.3.

2.3.5 Structure seule

Cette méthode permet de mettre en évidence les contours entre éléments de structure seulement. Elle se comporte exactement de la même manière que le rendu précédent

(*contour amélioré*), mais n'est pas sensible aux objets qui sont présents dans la scène. Ainsi, elle permet de récupérer uniquement les arêtes joignant les bâtiments entre eux et avec la route. La FIGURE 2.19 illustre ce procédé.

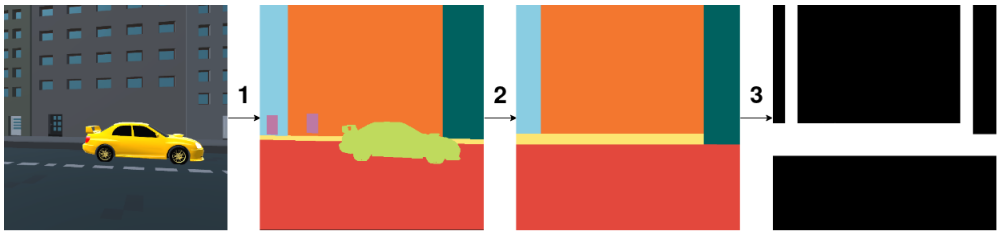


FIGURE 2.19 – Illustration de la méthode d'*extraction de structure seule*. Les étapes de construction sont les suivantes : (1) – Calcul de la carte de segmentation, soit grâce à des réseaux de neurones, soit en utilisant le mécanisme d'étiquettes de notre simulateur. Ici, les classes de segmentation sont propres à chaque entité de la scène. (2) – Mise à l'échelle de l'implant simulé (méthode analogue à *Scoreboard*, 2.3.1). Les objets comme la voiture ou les portes sont filtrés et n'apparaissent pas. (3) – Application des gradients (horizontal et vertical) pour détecter les changement d'objets.

Cette méthode est très efficace pour mettre en évidence les informations structurelles de la scène. Il faut encore une fois mettre en avant que cela n'est possible en temps réel qu'avec l'utilisation de réseaux de neurones.

2.3.6 Méthode combinée

Cette méthode est inspirée de ce qu'ont proposé Sanchez-Garcia et al. [Sanchez-Garcia et al., 2020]. Cette approche a pour but de combiner des informations de structure seule et les informations de segmentation sémantique d'objets. Il s'agit en fait d'une combinaison du rendu *structure seule* 2.3.5 et du rendu *extraction sémantique d'objets* 2.3.3. Pour combiner les deux informations, on regarde pour chaque pixel s'il y a une information sur la présence d'un objet dans la carte de segmentation sémantique (intensité du pixel non nulle). Si c'est le cas alors le nouveau pixel dans l'image finale sera de la même intensité. Sinon on attribue la valeur d'intensité de la carte de structure à ce pixel. La FIGURE 2.20 illustre ce procédé.

Ce procédé a l'avantage d'assembler les différentes caractéristiques détectables par chacune des deux composantes. Une restitution en temps réel repose sur la capacité à prédire à la fois une carte de segmentation sémantique et une carte de structure.

Nous venons de présenter les principales méthodes d'extraction d'informations présente dans notre simulateur. L'utilisation de ces méthodes, combinées avec le module de construction de représentation phosphénique présenté en 2.2, permet de créer une représentation phosphénique enrichie de la scène. Tous les calculs sont réalisés grâce au *pipeline*

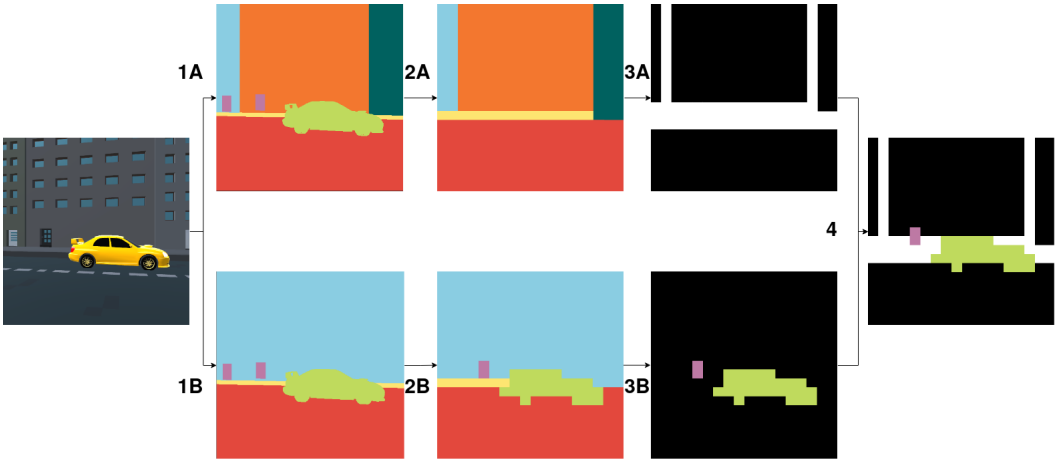


FIGURE 2.20 – Illustration de la méthode *combinée*. Les étapes de construction sont les suivantes : (1A) – Calcul de la carte de segmentation sémantique d'instances, (1B) - Calcul de la carte de segmentation sémantique classique. (2A) – Filtrage des objets et mise à la résolution de l'implant, (2B) – Mise à la résolution de l'implant. (3A) – Application des gradients pour extraire la structure, (3B) – Filtrage de la structure. (4) – Assemblage de 3A et 3B en image combinée.

graphique présent dans Unity 3D. Enfin, dans la section suivante (2.4), nous présentons le fonctionnement général du simulateur et ses principaux composants.

2.4 Fonctionnement du simulateur

Dans cette partie, nous présentons la manière dont le simulateur fonctionne un peu plus dans le détail. Nous expliquerons le fonctionnement du système de caméras mis en place (2.4.1), la manière dont les paramètres de restitution phosphéniques sont modifiables (2.4.2) et enfin, nous donnerons un exemple de restitution temps réel proposé par le simulateur (2.4.3).

2.4.1 Système de caméra

Nous décrivons ici le procédé utilisé pour construire les rendus phosphéniques associés à chaque méthodologie d'extraction de caractéristiques. Le processus est le même quelle que soit la méthodologie considérée. Dans un premier temps, une caméra est placée dans la scène 3D afin de réaliser une capture vidéo des éléments se trouvant face à elle. Cette caméra effectue le rendu de son image dans un composant de type *RenderTexture*. Ce composant est conçu pour recevoir le flux vidéo d'une caméra dans l'objectif de réaliser un post-traitement directement sur le composant. La FIGURE 2.21 illustre ces différents éléments.

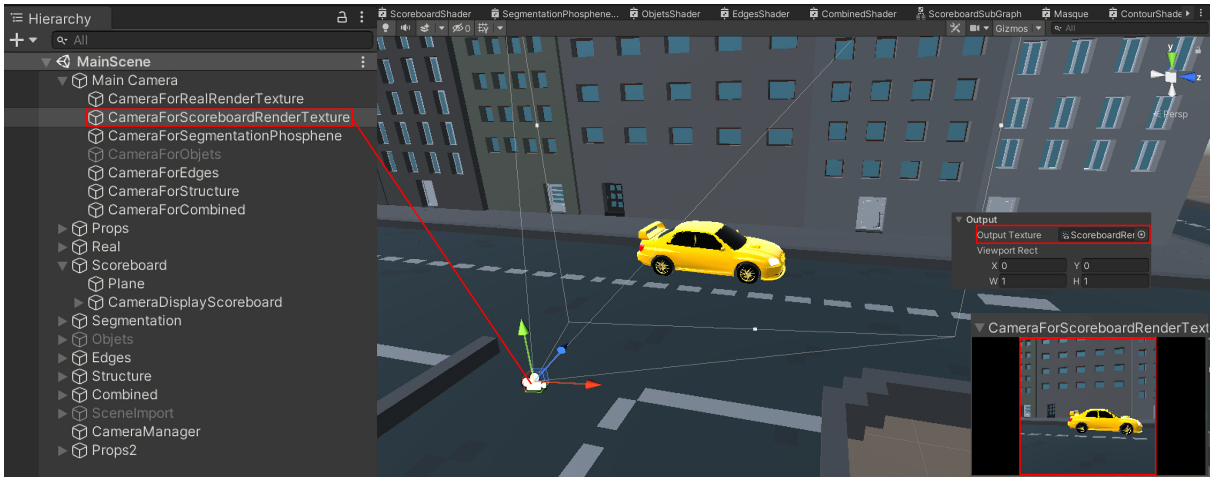


FIGURE 2.21 – Système de caméra. On retrouve ici la caméra dans l'onglet de gauche (CameraForScoreboardRenderTexture) qui effectue une capture de la scène (visible dans la fenêtre en bas à droite) et un rendu dans un RenderTexture (dans la section Output des paramètres de la caméra : ScoreboardRenderTexture)

Une fois que le rendu est effectué dans le *RenderTexture*, on peut alors l'associer à un plan de rendu et attacher un *shader* (processus de post-traitement graphique) à ce plan de rendu. C'est dans ce *shader* que l'on va spécifier le traitement de l'image ainsi que les paramètres associés. Une autre caméra réalise alors la capture du plan contenant l'image post-traitée et envoie ce flux vidéo dans la zone d'affichage principale. La FIGURE 2.22 montre ces différents éléments.

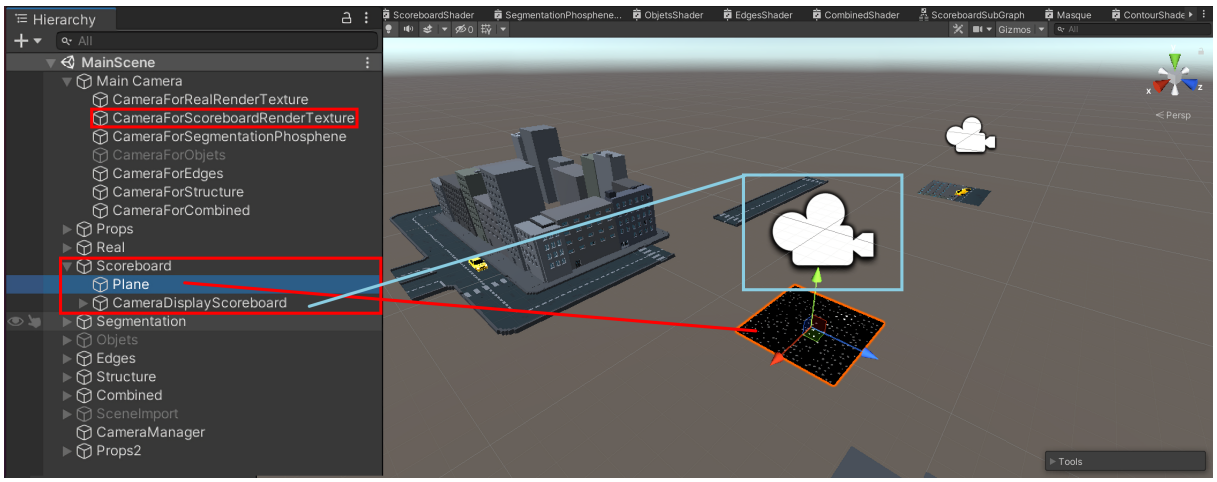


FIGURE 2.22 – Système de caméra, shader. Ici, le plan de rendu (Plane) reçoit le flux vidéo de la caméra de la FIGURE 2.21 et un shader applique un post-traitement. La caméra encadrée en bleu réalise l'acquisition du flux vidéo post-traité et l'affiche dans la zone de rendu principale (c'est-à-dire à l'utilisateur).

2.4.2 Sélection des paramètres

Pour régler à la fois les paramètres de chaque méthodologie de restitution et les paramètres de l'implant simulé, un menu de sélection est disponible directement depuis les paramètres du shader associé au plan de rendu. La FIGURE 2.23 montre le menu de sélection pour la méthode de restitution *scoreboard*.

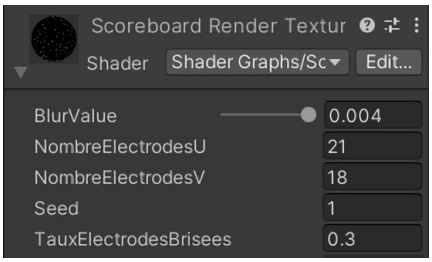


FIGURE 2.23 – Fenêtre de paramétrage du shader associé à la méthodologie scoreboard. On peut gérer depuis ce menu : le flou gaussien (puissance du flou), la résolution de l'implant et le taux d'électrodes brisées. Le paramètre "Seed" permet de générer une cartographie rétinotopique par valeur. Cela permet de simuler un implant différent pour chaque sujet lors des expériences. On peut également choisir le nombre de niveaux de gris utilisables par l'implant simulé.

2.4.3 Comparaison des méthodes de restitutions

Notre simulateur permet de réaliser de manière parallèle la restitution visuelle de toutes les méthodes présentées en 2.3. La FIGURE 2.24 montre les résultats des méthodologies appliquées dans deux contextes : intérieur et extérieur.

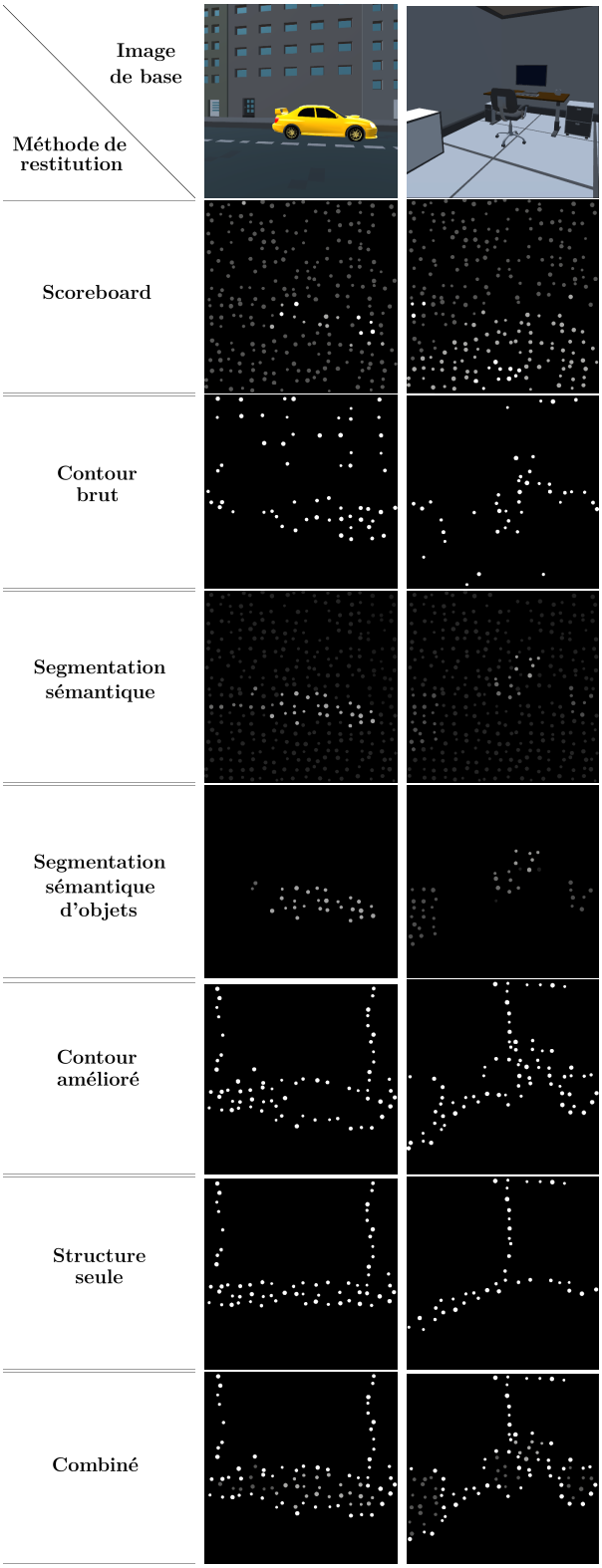


FIGURE 2.24 – Illustration des restitution phosphéniques selon les méthodologies présentées en 2.3. Deux environnements (extérieur et intérieur) sont présentés. Les paramètres de simulation sont les suivants : résolution de l'implant : 21x18, nombre de niveaux de gris : 4, taux d'électrodes brisées : 10%.

En résumé

Unity 3D est le moteur de simulation que nous avons choisi d'utiliser du fait de sa facilité de manipulation et de la quantité de ressources déjà disponibles (notamment des modèles 3D). Nous avons donc pu concevoir un prototype de simulation de vision prothétique en nous basant sur cet outil.

Après avoir présenté Unity 3D, nous avons ensuite détaillé la façon dont nous simulons les implants. Nous avons expliqué comment nous avons construit la représentation phosphénique qui permet de prendre en compte plusieurs phénomènes (problèmes) liés à l'utilisation d'une neuroprothèse visuelle. L'expérimentateur peut régler les paramètres suivants : la résolution de l'implant, le taux d'électrodes brisées, la cartographie rétinotopique, le nombre de niveaux de gris et la présence d'un flou gaussien.

Ensuite, nous avons expliqué comment extraire les caractéristiques des images et comment construire nos rendus. Le choix des caractéristiques à extraire est basé sur l'analyse des besoins des non-voyants issues du rapport de Ratelle et Couturier [Ratelle, 2019]. Nous avons ensuite expliqué comment utiliser le simulateur.

Les deux prochaines parties sont consacrées aux études que nous avons menées en utilisant le simulateur développé. La première étude (3) traite de l'impact de la présence d'indices sonores dans des tâches de navigation en intérieur sur la compréhension et la mobilité dans la scène. La seconde étude (4) compare deux méthodes de rendu à une troisième qui introduit un aspect interactif ou le sujet peut librement choisir le mode de restitution en temps réel.

Chapitre 3

Des indices audio pour améliorer la navigation et la compréhension avec une neuroprothèse

Sommaire

3.1 Utilisation de systèmes sonores dans l'aide à la navigation et la compréhension 81
3.1.1 Systèmes basés lunettes . . . 81
3.1.2 Systèmes basés canne . . . 84
3.2 Méthode et Matériel . . . 87
3.2.1 Simulateur de vision prothétique . . . 87
3.2.2 Rendus phosphéniques . . . 89
3.2.3 Hypothèses . . . 89
3.2.4 Sujets et conditions expérimentales . . . 90
3.3 Protocole . . . 91
3.3.1 Formulaire et présentation . . . 91
3.3.2 Familiarisation avec le dispositif expérimental . . . 91
3.3.3 Bloc d'expérience . . . 93
3.3.4 Évaluation finale . . . 95
3.4 Résultats . . . 97
3.4.1 Hypothèse 1 : Amélioration du comportement de navigation des sujets 97
3.4.2 Hypothèse 2 : Amélioration des capacités d'analyse et de perception du sujet . . . 97
3.4.3 Hypothèse 3 : Préférences de l'utilisateur . . . 100
3.5 Analyse des résultats et discussion . . . 102
3.5.1 Comportement des sujets . . . 102

3.5.2 Capacités d'analyse et de perception du sujet 102
3.5.3 Satisfaction et classement 102
3.5.4 Concernant les différents type de rendus 102
3.5.5 Concernant la simulation 103
3.5.6 Concernant les retours audio 103

Dans les chapitre précédents, nous avons présenté les différents travaux réalisés concernant la restitution visuelle dans un implant simulé. Dans ce chapitre, nous présentons la première étude de vision prothétique simulée que nous avons conduite qui traite de l'intérêt d'augmenter la restitution visuelle prothétique par des indices audio décrivant la scène.

En effet, la majorité des systèmes de neuroprothèses visuelles reposent uniquement sur la communication d'informations par la voie visuelle. Notre réflexion nous a amené à considérer la possibilité de donner des indices audio simples pendant l'utilisation d'une neuroprothèse visuelle afin de désambiguïser l'environnement et d'améliorer la compréhension dans un contexte de navigation. Ce choix est basé sur les besoins exprimés par les non-voyants dans le rapport de Ratelle et Couturier [Ratelle, 2019].

Afin d'évaluer l'intérêt de l'ajout de tels indices, nous avons mis en place une étude utilisateur. Nous montrons que l'ajout d'indices audio qui identifient l'objet apparaissant au centre du champ de vision permet d'améliorer de manière significative la compréhension de la scène, par rapport à l'utilisation de la restitution phosphénique seule. L'apport des indices audio pour la compréhension et la navigation est également bien supérieur à celui apporté par une amélioration du rendu visuel.

Nous présentons dans un premier temps quelques systèmes utilisant l'audio comme outil informatif contextuel pour les non-voyants (3.1). Ensuite, nous détaillons plus précisément l'expérience mise en place, avec une description des méthodes et du matériel utilisé (3.2) et l'explication du protocole expérimental (3.3). Enfin, nous présentons nos résultats (3.4) et les analysons (3.5).

3.1 Utilisation de systèmes sonores dans l'aide à la navigation et la compréhension

Dans cette section, nous présentons les différents dispositifs conçus et utilisés à notre connaissance pour aider les non-voyants à appréhender leur environnement. Ces systèmes se décomposent en trois grandes catégories : les systèmes basés canne, les systèmes basés lunettes et les autres systèmes.

3.1.1 Systèmes basés lunettes

Dans cette partie, nous présentons des systèmes audio d'aide aux non-voyants dans différents types de tâches : détection d'obstacles, lecture de texte et navigation.

Détection d'obstacles

La première utilisation de ce type de dispositif est la détection et la communication d'obstacles. Ces dispositifs utilisent généralement un capteur émettant un flux d'ultrasons, et qui ensuite détermine la distance aux objets à proximité. Une fois la distance établie, un signal sonore de durée et volume variable est émis.

[Sadi et al., 2014] proposent un dispositif monté sur des lunettes équipées d'un capteur à ultrasons tel que présenté dans la FIGURE 3.1. Sur la FIGURE 3.2, on retrouve les propriétés de communications sonores basées sur la distance par rapport à l'objet dans le cadre de l'étude de Sadi et al.



FIGURE 3.1 – Le prototype de restitution sonore d'informations de proximité aux objets de Sadi et al. Il est composé d'un capteur ultrasons, d'une alarme et d'un boîtier de calcul. [Sadi et al., 2014].

Distance de l'obstacle (m)	Durée du signal (s)	Intervalle de temps (s)	Volume du signal (dB)
Entre 2 et 3 mètres	1 seconde	1 seconde	40
Entre 1.5 et 2 mètres	0.8 seconde	0.8 seconde	45
Entre 1 et 2 mètres	0.5 seconde	0.5 seconde	50
Moins de 1 mètre	0.3 seconde	0.3 seconde	55

FIGURE 3.2 – Ce tableau récapitule les principales caractéristiques d'émission de sons dans le dispositif proposé par Sadi et al. On y retrouve les seuils de distances, les durées effectives de signaux audio, les intervalles de temps et le volume du son [Sadi et al., 2014].

Des dispositifs similaires ont été proposés par [Kassim et al., 2016], avec en plus un retour sous forme de vibration. Le dispositif sonore est montré sur la FIGURE 3.3 et le dispositif vibratoire sur la FIGURE 3.4.

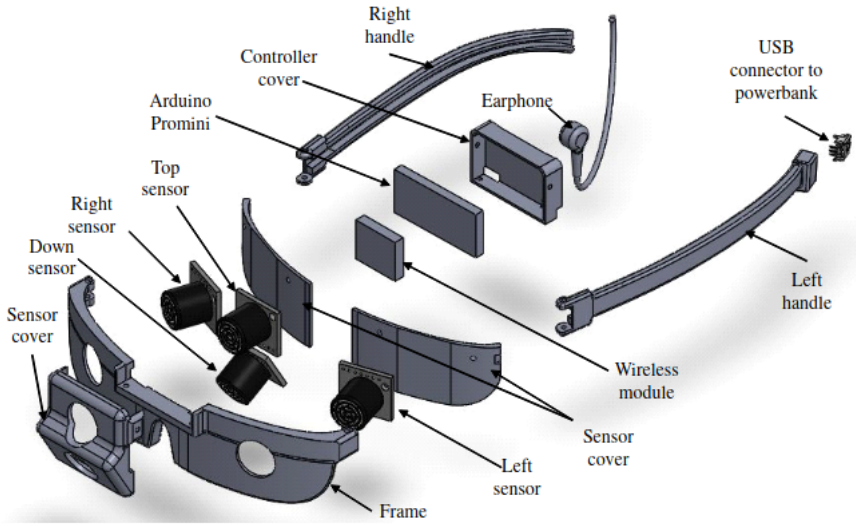


FIGURE 3.3 – Le prototype de restitution sonore d'informations de proximité aux objets de Kassim et al. Il est composé de plusieurs capteurs sensitifs [Kassim et al., 2016].

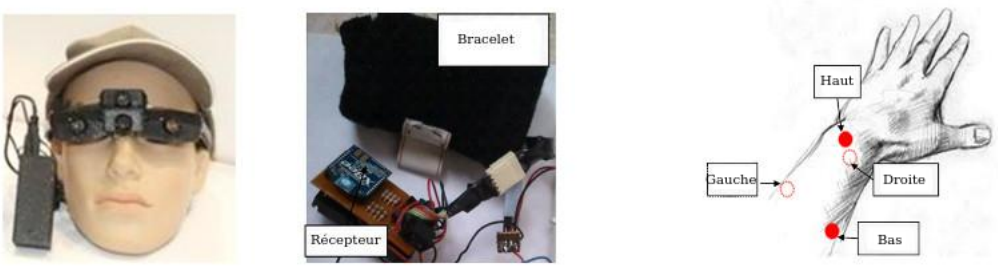


FIGURE 3.4 – Le prototype de restitution par vibration de Kassim et al. De gauche à droite, le dispositif sonore, la bracelet vibratoire et les zones de stimulation sur le bras [Kassim et al., 2016].

Lecture de texte

Une autre possibilité d'utilisation est la lecture de texte depuis une source naturelle. Dans ce secteur d'utilisation, une caméra RGB scanne la zone devant le non-voyant, réalise une détection de texte suivie d'une conversion en son via une synthèse vocale. [Yi and Tian, 2015] proposent un système réalisant ce procédé, comme illustré sur la FIGURE 3.5.

Navigation

Dans un contexte de navigation, plusieurs études ont été réalisées. Par exemple [Wang et al., 2013b] se sont concentrés sur l'intérêt de mettre en évidence les panneaux et les portes. Ils proposent de faire en sorte que le système puisse indiquer à l'utilisateur s'il se trouve en face des commodités pour hommes ou pour femmes. Le système détecte dans un premier temps la présence d'une porte, puis analyse le panneau qui se trouve dessus

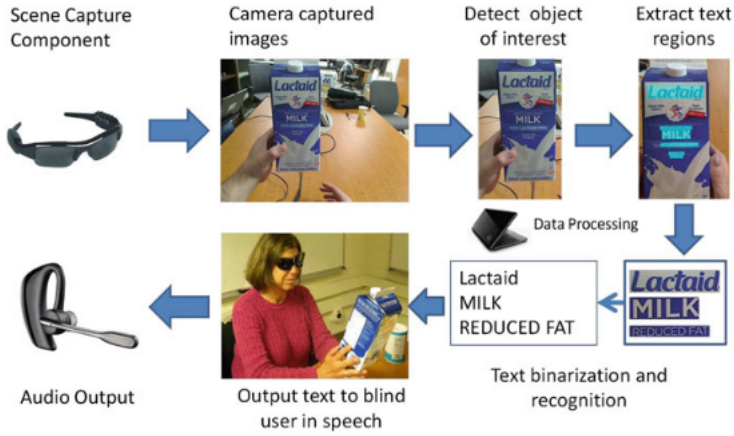


FIGURE 3.5 – Illustration du "pipeline" de conversion texte vers parole proposé par [Yi and Tian, 2015]. Les principales étapes sont la détection d'objets, l'extraction des zones de texte puis la transformation en texte simplifié afin de pouvoir faire un retour audio.

pour ensuite restituer une information à l'utilisateur via un retour audio.

Une autre étude réalisée par [Stoll et al., 2015] propose d'utiliser une caméra de profondeur. L'objectif est de générer un retour audio à partir des informations traitées et extraites des images de profondeur. Dans leur étude, les participants avaient les yeux bandés. Ils n'ont donc jamais vu le dispositif expérimental. Après avoir été conduits à un point de départ, à l'aide de leur dispositif, on leur a demandé de marcher jusqu'à la fin d'un chemin. Il leur était demandé de naviguer aussi rapidement et précisément que possible, c'est-à-dire en minimisant les contacts avec les murs. Avant de commencer les essais, ils étaient informés qu'il n'y avait pas de barrières, d'escaliers, de culs-de-sac ou de couloirs étroits. Pendant les essais, les participants étaient avertis lorsqu'ils commençaient à aller dans la mauvaise direction (c'est-à-dire à faire demi-tour). La FIGURE 3.6 illustre leur dispositif.

3.1.2 Systèmes basés canne

L'utilisation d'une canne d'assistance est essentielle pour réduire le risque de collision, ce qui peut aider les personnes en situation de handicap visuel à marcher avec plus de confiance. En général, une canne d'assistance est développée en montant des modules de détection et de rétroaction sur une canne blanche classique. Par la suite, la canne d'assistance acquiert des informations sur l'environnement et transmet un retour aux utilisateurs.

Navigation

De nombreux travaux ont été réalisés dans le but d'ajouter un dispositif d'aide sur des cannes dans un contexte de navigation. On recense trois grandes familles de méthodologie

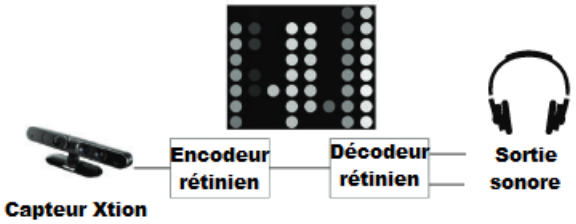


FIGURE 3.6 – Illustration issue de [Stoll et al., 2015]. Le capteur Xtion réalise l'acquisition de la scène devant le sujet. Un algorithme extrait les informations nécessaires à l'établissement d'une version simplifiée de la scène (semblable à une restitution phosphénique, avec un implant de résolution 8x8). Une fois cette version simplifiée construite, des signaux audios sont envoyés dans un casque porté par le sujet pour le guider sur le parcours.

d'acquisition : des capteurs ultrasons, des capteurs GPS et des capteurs RGB-D (red green blue - depth / rouge vert bleu - profondeur) [Fan et al., 2014, Gupta et al., 2015].

Fan et al. [Fan et al., 2014], ont utilisé une caméra RGB-D et un capteur à ultrasons pour acquérir des informations visuelles dynamiques sur l'environnement et détecter les obstacles environnants. La caméra RGB-D est capable d'obtenir des vidéos synchronisées à la fois en couleur et en profondeur. Pour mettre en œuvre la navigation en extérieur, ils ont ajouté un module GPS. Les résultats des expériences de validation menées dans une zone ouverte ont démontré que la canne d'assistance installée dans cette unité de détection peut aider les personnes visuellement handicapées à se déplacer à l'extérieur en toute sécurité. Cependant, cette canne ne peut pas traiter les données d'image capturées par la caméra RGB-D en temps réel. La FIGURE 3.7 montre le prototype de canne proposé par [Fan et al., 2014].

Gupta et al. [Gupta et al., 2015], ont utilisé un capteur à ultrasons et un récepteur GPS dans des cannes classiques. L'ajout d'un module GPS permet aux personnes visuellement handicapées de se déplacer à l'extérieur grâce à un réseau de satellites. Les signaux audio sont générés par Pygame, un module de programmation permettant de créer des jeux et des animations.

Détection d'obstacles

La détection d'obstacles via cannes est également un sujet qui a été largement abordé [Shim and Yoon, 2002, Scherlen et al., 2007, Kim et al., 2009, Kumar et al., 2014, Dias et al., 2015, Sharma et al., 2017, Wankhade et al., 2017]. La majorité des systèmes de ce type se basent sur des capteurs ultrasons, des lasers ou encore des caméras RGB-D.

Kumar et al. [Kumar et al., 2014] ont développé une canne à ultrasons pour aider les aveugles à se déplacer. Elle est équipée de trois paires d'émetteurs-récepteurs d'ultrasons, ce qui permet aux aveugles de percevoir les obstacles aériens et terrestres ainsi que les nids



FIGURE 3.7 – Canne utilisée lors de l'expérience de Fan et al. Elle est composée d'une caméra RGB-D et d'un capteur à ultrasons[Fan et al., 2014].

de poule qui se trouvent devant eux grâce à des avertissements sonores.

Kim et al. [Kim et al., 2009] ont utilisé leur canne intelligente pour mesurer la luminosité de l'environnement. Pour détecter avec précision les obstacles situés à l'avant, Shim et Yoon [Shim and Yoon, 2002] ont fixé sur leur canne deux antennes utilisées comme capteurs de contact, un capteur à ultrasons et un capteur infrarouge. Grâce à l'ajout des capteurs de contact, la canne est très efficace pour la détection des obstacles à courte distance.

Autres systèmes communiquant un retour d'information audio

Des systèmes très variés pour communiquer de l'information visuelle sous forme audio ont vu le jour en plus des systèmes directement intégrés sur la canne ou sur des lunettes. La plupart de ces systèmes sont fixés sur des habits. On retrouve notamment des systèmes utilisant des chapeaux, des ceintures ou des vestes [Mahalle, 2014, Mocanu et al., 2016, Srinivas et al., 2020].

Dans la suite, nous expliquons comment nous avons intégré un système d'indices audio dans notre simulateur, et nous détaillons plus précisément le protocole de l'expérience que nous avons menée.

3.2 Méthode et Matériel

Dans cette partie, nous présenterons les particularités de la version du simulateur (présenté au chapitre 2) utilisée dans cette étude. Puis nous présenterons successivement les méthodes de restitution utilisées, les conditions expérimentales choisies et la tâche expérimentale.

3.2.1 Simulateur de vision prothétique

Pour conduire cette étude, nous avons utilisé le simulateur présenté dans le chapitre 2 pour générer les différents rendus. La spécificité de cette étude est que les sujets doivent réaliser une tâche de mobilité (déplacement dans un environnement 3D). Après la conception du dispositif expérimental, un logiciel (windows/linux) est rendu accessible. Un numéro de passage est donné à chaque sujet expérimental et les données sont récupérées par courrier électronique sur une adresse conçue à cet effet. Les données sont ensuite récupérées et enregistrées sur un serveur privé.

Modèles utilisés

Les scènes de simulation que nous utilisons sont composées de 4 pièces interconnectées. Chaque pièce contient deux objets que l'on trouve généralement dans un appartement (par exemple un canapé, une table, ou encore des chaises). Un exemple de scène utilisée est illustré sur la Figure 3.8.

Scènes modélisées

Nous avons modélisé quatre scènes d'intérieur de topologies variées. Les quatre scènes sont schématisées et visibles sur la Figure 3.9.

Implant simulé

Nous avons choisi de simuler le système PRIMA développé par Pixium Vision (Paris, France), composé d'un réseau de 21x18 électrodes. Notre choix est motivé par le fait que c'est l'implant le plus récent ayant passé des phases d'essais cliniques satisfaisantes [Fernandez, 2018]. Il est aussi l'un de ceux qui permet un rendu visuel avec la plus grande résolution à ce jour. En effet, les limitations techniques actuelles ne permettent pas d'avoir un nombre d'électrodes très élevé (l'Argus II possède seulement 6x10 électrodes [Farvardin et al., 2018]).

L'image est capturée avec un champ visuel de 50°x50°. Un taux d'électrodes défectueuses de 10% est appliqué, afin de simuler des électrodes non-fonctionnelles ou cassées. Un flou gaussien est appliqué aux phosphènes générés, la taille des phosphènes d'un même implant variant entre 0,34° et 0,67° du champ de vision. L'espacement entre deux phosphènes varie

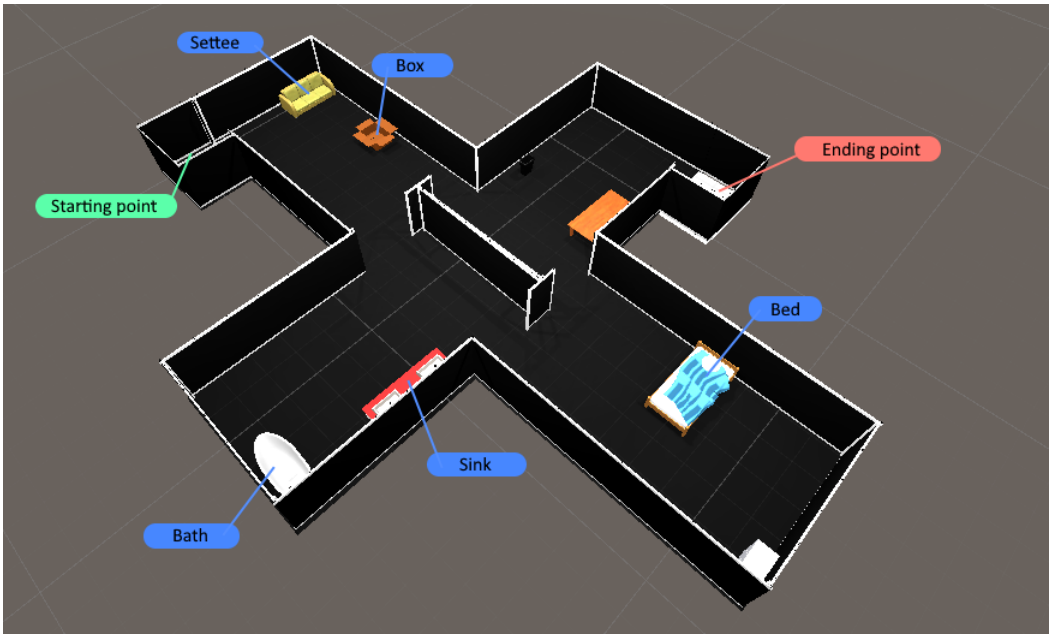


FIGURE 3.8 – Vue d'ensemble d'une des scènes utilisées dans nos expériences. L'utilisateur doit naviguer du point de départ au point d'arrivée et visiter les quatre différentes pièces qui contiennent divers objets : chambre (lit, commode), bureau (table, poubelle), salon (carton, canapé) et salle de bain (baignoire, lavabo).

entre 0,67° et 2,7° du champ de vision. Chaque électrode génère un phosphène avec quatre niveaux de gris.

Sur la base de ces règles, une disposition différente des phosphènes est générée aléatoirement pour chaque sujet (grâce au mécanisme de *seed* présenté dans le chapitre 2), afin de modéliser le scénario du monde réel où deux personnes implantées avec le même modèle de prothèse virtuelle peuvent avoir des perceptions phosphéniques différentes.

Utilisation pratique

Pour naviguer dans l'environnement, les sujets utilisent le clavier. La touche "espace" permet d'avancer et les touches directionnelles droite et gauche permettent de tourner la caméra à droite et à gauche. Lorsque les sujets se déplacent, un clip audio de pas est joué à chaque déplacement. Un autre clip est joué lorsque le sujet heurte un obstacle (mur ou objet). En plus des bruits de pas et de collision, le nom des objets est prononcé lorsqu'ils apparaissent dans la rangée centrale de l'écran (fenêtre verticale qui couvre un tiers de l'écran, centrée au milieu). De plus, nous ne nommons qu'un seul objet détecté à la fois pour deux raisons, nous voulons que la détection soit consciente, c'est-à-dire que le sujet doive fixer un objet spécifique pendant 2 secondes pour qu'il soit nommé, et d'autre part pour éviter une surcharge d'informations cognitives si plusieurs objets sont détectés.



FIGURE 3.9 – Schéma des quatre scènes choisies dans le cadre de cette étude. On retrouve à chaque fois les quatre pièces modélisées et meublées d'objets répartis aléatoirement à l'intérieur (salon, salle de bain, chambre, bureau). Les cellules d'entrées et de sorties sont représentées en gris. Les rectangles bleu foncé représentent les portes qui permettent de passer d'une pièce à une autre.

3.2.2 Rendus phosphéniques

Dans cette étude, nous avons utilisé le rendu *Scoreboard* comme condition de contrôle (2.3.1), utilisé par défaut dans certains implants commercialisés actuellement. Nous avons aussi utilisé un rendu mélangeant les segmentations d'objets et de structure (appelé *Combiné*) récemment proposé par Sanchez et al. [Sanchez-Garcia et al., 2020] (2.3.6). La méthodologie de construction de ces différents rendus est abordée dans le chapitre 2.

3.2.3 Hypothèses

Nos hypothèses pour cette étude sont les suivantes :

L'ajout d'indices sonores couplé à un mode de rendu visuel sélectif (objets + structure) devrait améliorer à la fois :

— le comportement du sujet dans la scène (diminution du temps de navigation et diminution du nombre de collisions avec les obstacles sur le chemin),

— la perception du sujet (reconnaissance de la scène, reconnaissance et localisation des objets).

Nous faisons également l'hypothèse que les sujets préféreront ce mode de rendu (objets + structure + audio).

3.2.4 Sujets et conditions expérimentales

23 sujets âgés de 11 à 55 ans (moyenne = 25, écart-type = 10) ont participé à l'étude. Ces sujets ont été recrutés d'une part dans une école d'ingénieur et d'autre part via les réseaux sociaux (LinkedIn, Facebook). Chaque participant avait une vision normale ou corrigée. Tous les sujets ont donné leur consentement pour participer à l'étude et permettre le stockage de leurs données anonymisées, en accord avec le RGPD.

Nous avons quatre conditions expérimentales dans cette étude :

— Contrôle (*Scoreboard* – 2.3.1) ;
— Contrôle + audio ;
— *Combiné* (structure + objets) – 2.3.6 ;
— *Combiné* + audio.

Tâche et variable dépendantes

Les sujets doivent naviguer dans 4 environnements différents (un pour chaque condition) du point de départ au point d'arrivée le plus rapidement possible, tout en identifiant et en localisant le plus d'objets possible dans la scène. À la fin de chaque session, il est demandé aux utilisateurs de reconnaître la disposition des pièces parmi un ensemble de 4 possibilités, puis de localiser sur la bonne carte de la scène les objets qu'ils pensent avoir identifiés. Ils sont ensuite invités à remplir un questionnaire pour chaque condition expérimentale. L'ensemble de l'expérience dure environ une heure.

Afin d'évaluer la facilité d'utilisation d'un rendu, nous mesurons 4 paramètres :

— la mobilité dans la scène en comptant le nombre de collisions avec les objets et en mesurant le temps total pour parcourir l'itinéraire ;
— la perception de la scène en comptant le nombre d'objets correctement perçus (c'est-à-dire le nombre d'objets identifiés qui se trouvent réellement dans la scène) ;
— la compréhension de la scène, en comptant le nombre d'objets correctement localisés sur la carte après la navigation (c'est-à-dire le nombre d'objets identifiés qui sont dans la scène et qui sont également bien placés sur la carte de la scène) ;
— la satisfaction de l'utilisateur pour chaque rendu en demandant à l'utilisateur de remplir un questionnaire MeCue standard [Minge et al., 2017]. Le questionnaire MeCue utilisé est disponible en annexe.

Pour évaluer la compréhension de la scène, on calcule un score de placement pour chaque couple (sujet – condition). Pour ce faire, on découpe chacune des quatre pièces composant

la scène en six cellules identiques. Un objet bien labellisé et placé dans la bonne cellule augmente le score de 1. Si l'objet est à cheval sur au moins deux cellules dont l'une des deux est la bonne, alors le score augmente de 0,5. Dans tous les autres cas, l'objet est considéré comme mal placé. Des exemples de placement d'objets réalisés par les sujets sont montrés sur le Tableau 3.1.

Pour récupérer les trajets des sujets pendant la navigation, on enregistre leur position dans la scène 3D à chaque fois qu'il y a une interaction de leur part. Nous pouvons ainsi rejouer leurs déplacements. Le Tableau 3.2 montre deux exemples de trajectoires effectuées par les sujets pendant l'expérience.

Les données récupérées prennent la forme suivante :

— un nombre entier décrivant le nombre de collisions et un nombre réel pour la durée du trajet ;
— un nombre entier entre 0 et 8 décrivant le nombre d'objets bien perçus et identifiés ;
— un nombre entier compris entre 0 et 8 décrivant le nombre d'objets bien perçus et bien localisés sur la carte de la scène ;
— les scores des réponses aux questions sur une échelle de Likert en 7 points allant de "tout à fait d'accord" à "pas du tout d'accord".

3.3 Protocole

3.3.1 Formulaire et présentation

Les sujets sont d'abord invités à remplir un formulaire (nom, prénom, âge, état de vision, expérience avec les jeux vidéo). Ils doivent également accepter les conditions de conservation des résultats et de leurs données personnelles (conformément au RGPD). Les sujets sont ensuite initiés à la tâche principale de l'expérience en lisant une explication des tâches à effectuer au cours de l'étude. Ils peuvent se familiariser avec les différents clips audio qui sont joués pendant la navigation (bruits de pas, collisions, rotation). Avant chaque bloc d'étude, les sujets peuvent faire une courte pause pour se reposer et se préparer au bloc suivant. Il est précisé qu'il est impératif de ne pas s'arrêter une fois qu'un bloc a commencé pour ne pas fausser les résultats, notamment les temps de parcours.

3.3.2 Familiarisation avec le dispositif expérimental

Dans la phase préliminaire, les sujets doivent naviguer dans une scène contenant quelques objets sans modification du rendu. Ils voient la scène telle qu'elle est réellement. Après la navigation, ils sont confrontés à une étape d'évaluation au cours de laquelle ils doivent sélectionner les objets qu'ils pensent avoir perçus pendant la navigation, choisir la carte

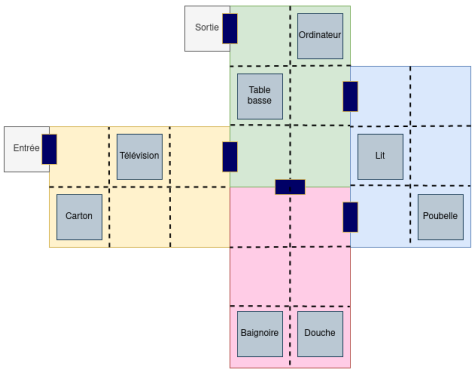
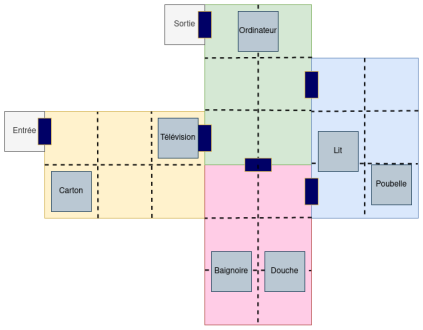
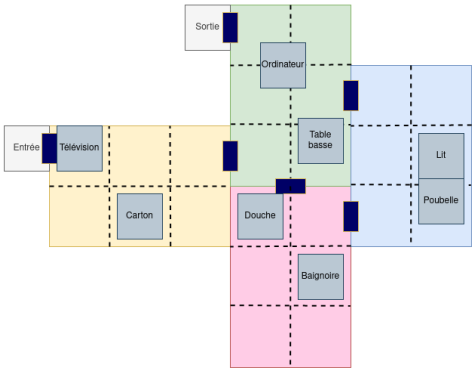
Description	Placement
Vérité terrain La scène est composée de 4 pièces contenant chacune 2 objets. Cette carte de placement correspond à une des scènes dans laquelle les sujets pouvaient naviguer.	
Carte proposée par le sujet 13 Cette carte de placement correspond à une des cartes fournie par le sujet numéro 13. Le sujet utilisait à ce moment là le mode de rendu combiné avec indices audios. Le calcul du score s'effectue comme suit : 0,5 point pour l'ordinateur + 1 point pour le carton + 0,5 point pour la douche + 0,5 point pour la baignoire + 0,5 point pour le lit + 1 point pour la poubelle. Ce qui donne un score de 4/8.	
Carte proposée par le sujet 21 Cette carte de placement correspond à une des cartes fournie par le sujet numéro 21. Le sujet utilisait à ce moment là le mode de rendu combiné avec indices audios. Le calcul du score s'effectue comme suit : 1 point pour la poubelle + 1 point pour l'ordinateur. Ce qui donne un score de 2/8.	

TABLEAU 3.1 – Exemple de placements d'objets réalisés par les sujets 13 et 21 comparés à la vérité terrain. Le calcul des scores de placement est également détaillé.

de la scène et enfin placer les objets sur cette carte.

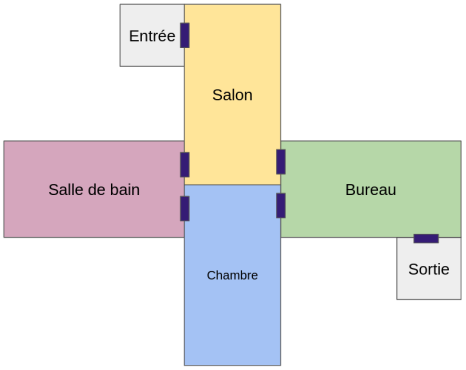
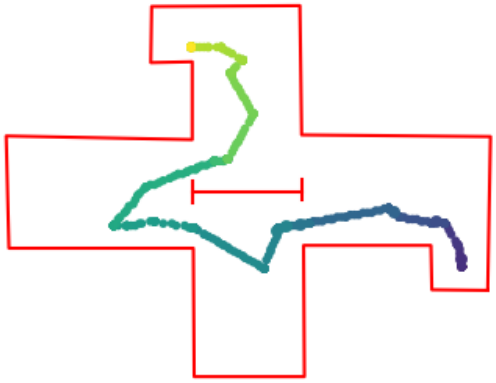
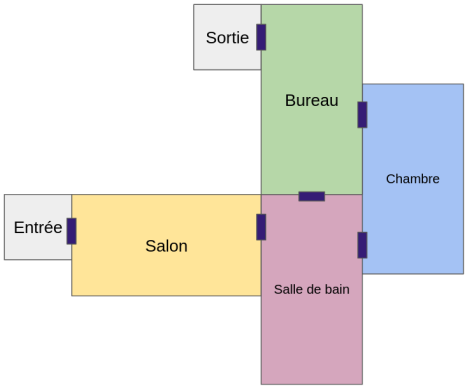
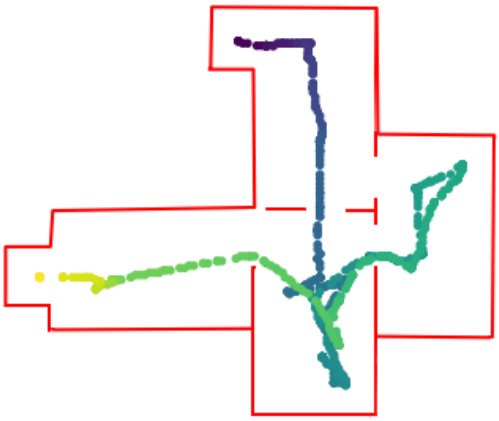
Scène	Trajectoire du sujet
	
	

TABLEAU 3.2 – Exemple de trajectoires suivies par les sujets pendant la phase d'expérimentation. Les positions successives sont enregistrées ce qui permet de voir les chemins suivis plus facilement. Le trajet commence par les points jaunes et termine par les points bleus.

3.3.3 Bloc d'expérience

Chaque bloc de l'expérience est décomposé en 3 parties distinctes : familiarisation, navigation, et évaluation.

Lors de l'étape de familiarisation, les sujets ne sont pas évalués. Ils doivent seulement naviguer dans une scène plus simple, comportant quelques objets afin d'être confrontés et de s'habituer au rendu qu'ils devront utiliser lors de l'étape de navigation.

Une fois cette étape de familiarisation terminée (atteindre une zone d'arrivée dans la scène), les consignes de navigation sont rappelées : parcourir la scène le plus rapidement possible, identifier la configuration des pièces de la scène, identifier un maximum d'objets afin de les placer sur un plan et de pouvoir les nommer. Pour faciliter cette étape, un chemin de navigation est proposé. Il est constitué d'une série d'objets uniques à identifier,

les conduisant à travers toutes les pièces de la scène, et se terminant dans la zone d'arrivée. Les sujets disposent d'un maximum de 12 minutes pour réaliser cette étape. Au-delà de cette durée, l'exercice prend fin et le sujet passe à la phase d'évaluation.

Enfin vient la phase d'évaluation. Il est demandé aux sujets de cocher d'abord les objets qu'ils pensent avoir rencontrés dans la scène (FIGURE 3.10), puis de sélectionner la topologie de la scène dans laquelle ils pensent avoir navigué (FIGURE 3.11). Il leur est ensuite demandé de positionner les objets précédemment cochés sur la bonne scène (même s'ils se sont trompés lors de la sélection de la scène, FIGURE 3.12). Enfin, ils remplissent le questionnaire (FIGURE 3.13) et se voient proposer une saisie en texte libre pour donner leur avis sur le bloc qu'ils viennent de terminer. Nous leur demandons des commentaires sur le rendu utilisé, les difficultés rencontrées, et de nous informer des problèmes techniques non liés à l'étude.

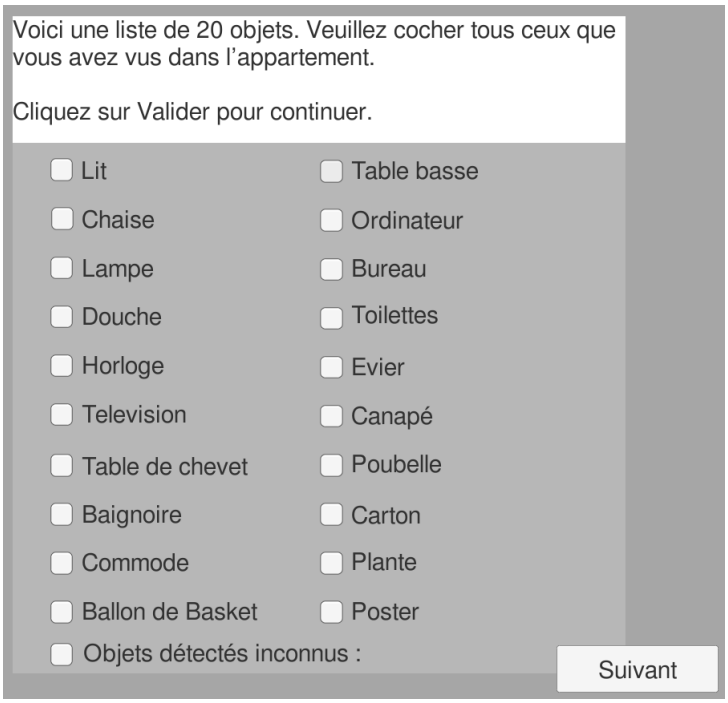


FIGURE 3.10 – Interface : sélection des objets perçus. Les sujets peuvent cocher autant d'objets qu'ils le souhaitent. Dans le cas où un objet serait perçu et non reconnu, ils peuvent cocher la case "Objets détectés inconnus". Ils devront alors fournir un nombre d'objets détectés inconnus.

Pour chaque sujet participant à l'étude, l'ordre d'apparition des conditions et des scènes est aléatoire. Cela permet d'éviter une partie des biais qui pourrait apparaître pendant l'expérience (accommodation au simulateur, navigation plus rapide et confiante en fin d'expérience, phénomène d'apprentissage).

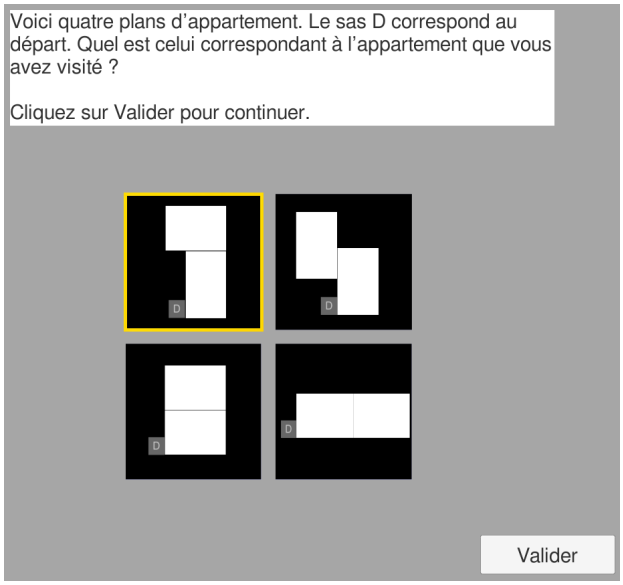


FIGURE 3.11 – Interface : sélection du plan de la scène. Il suffit aux sujets de cliquer sur une des quatre scènes affichées.

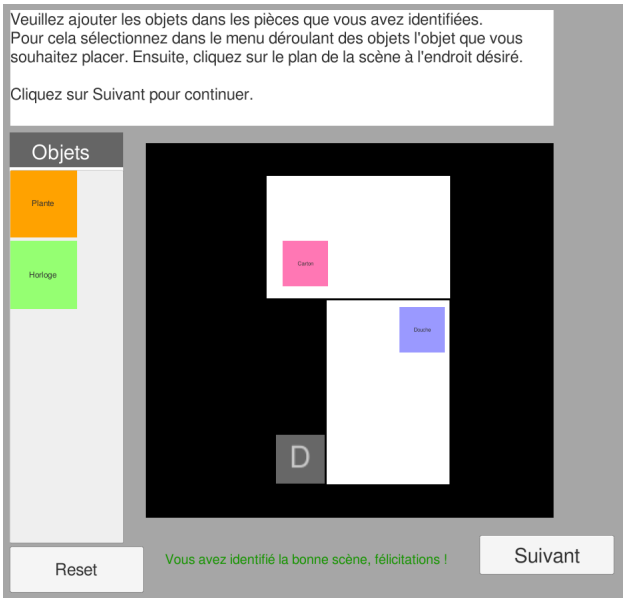


FIGURE 3.12 – Interface : placement des objets perçus dans la scène. Les sujets peuvent déplacer les objets à volonté avant de valider leur réponse. Ils peuvent passer la phase de sélection s'ils ont oubliés où se trouvent les objets.

3.3.4 Évaluation finale

À la fin des quatre blocs d'expérience, les sujets doivent classer les rendus de leur favori à leur moins favori (FIGURE 3.14). Enfin, ils peuvent faire un commentaire global sur l'expérience. Ces commentaires nous permettent de comprendre le ressenti des sujets et de pouvoir tenir compte d'éventuels problèmes techniques non liés à l'expérience en elle-même

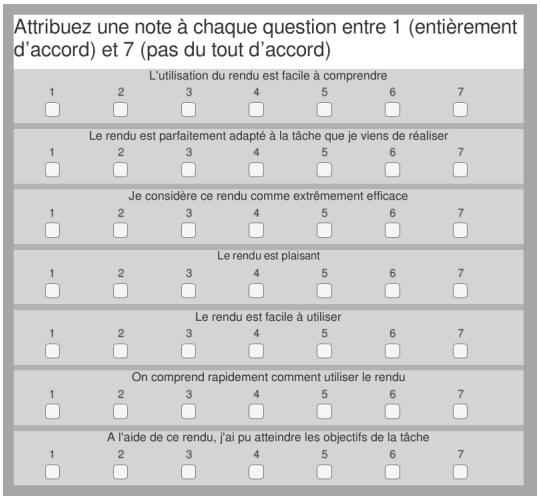


FIGURE 3.13 – Interface : questionnaire de satisfaction MeCue adapté à l'expérience. Les sujets doivent cliquer sur la "checkbox" qui correspond à leur ressenti. La liste des questions posées est disponible en annexe.

(coupure de courant par exemple).

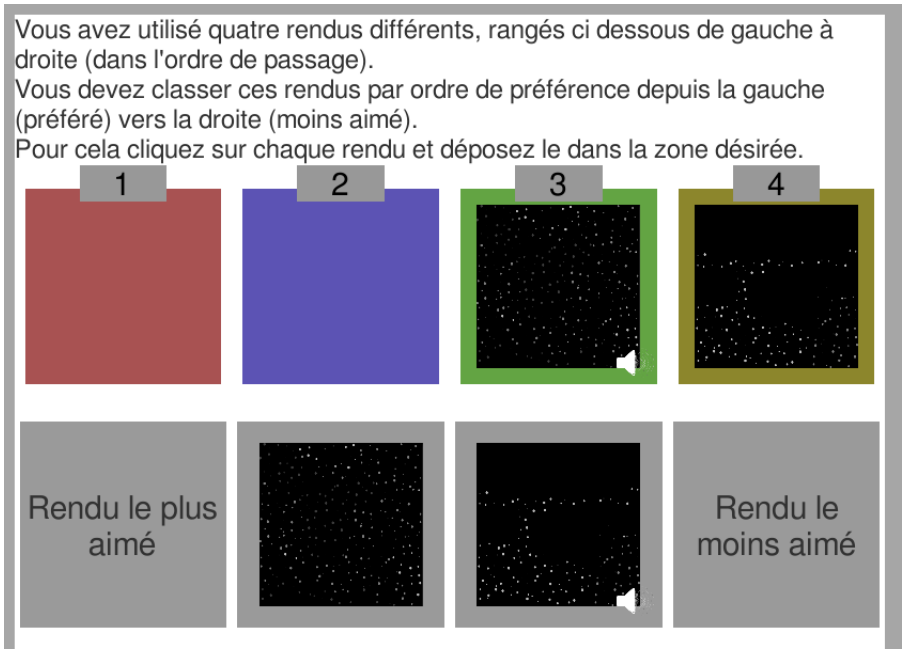


FIGURE 3.14 – Interface : classement des rendus. Les sujets classent les conditions de celle qu'ils ont la plus aimée à celle qu'ils ont la moins aimée. Pour faciliter la reconnaissance, une couleur est associée à chaque condition pendant toute la durée de l'expérience.

Nous venons de présenter le protocole expérimental et l'interface auxquels les sujets ont été soumis pendant l'expérience. Nous allons maintenant nous intéresser aux résultats constatés de l'expérience conduite.

3.4 Résultats

Pour analyser les résultats quantitatifs, nous avons effectué des tests ANOVA à 1 facteur selon le modèle (Condition), puis utilisé un test post-hoc de Tukey pour comparer les paires deux à deux. Dans les figures, nous avons également indiqué des intervalles de confiance à 95%.

Nous discuterons les résultats dans cet ordre : résultats liés à l'amélioration du comportement de navigation des sujets, résultats liés à l'amélioration des capacités d'analyse et de perception du sujet et enfin résultats sur les préférences utilisateurs. Dans cette partie, nous ne présenterons que les résultats, les analyses seront réalisées par la suite en section 3.5.

3.4.1 Hypothèse 1 : Amélioration du comportement de navigation des sujets

Temps de parcours

Le temps de déplacement est le temps total mis par le sujet pour effectuer un bloc de navigation. Ce temps est noté en secondes et est limité à 720 (le déplacement est arrêté s'il dure plus de 12 minutes). Comme les temps de parcours ne sont pas distribués selon une loi normale, nous effectuons le test ANOVA sur les données transformées par une opération logarithmique. Une ANOVA à un facteur (Condition) a montré que le temps de déplacement pendant les simulations n'était pas significativement différent selon les conditions ($F(3,84) = 0{,}284$; $p = 0{,}837$). La Figure 3.15 montre ces résultats.

Nombre de collisions

Le nombre de collisions correspond au nombre d'impacts avec les objets et les murs de la scène. Une ANOVA à un facteur (Condition) montre que le nombre de collisions n'est pas significativement différent selon les conditions ($F(3,79) = 0{,}574$; $p = 0{,}634$). La Figure 3.16 montre ces résultats.

3.4.2 Hypothèse 2 : Amélioration des capacités d'analyse et de perception du sujet

Reconnaissance des objets

La reconnaissance des objets est évaluée en comptant pour chaque rendu le nombre moyen d'objets identifiés par les sujets, qui sont réellement présents dans la scène. Une ANOVA à un seul facteur (Condition) a montré que le nombre d'objets bien reconnus lors des simulations est significativement différent selon les conditions ($F(3, 84) = 16.569$;

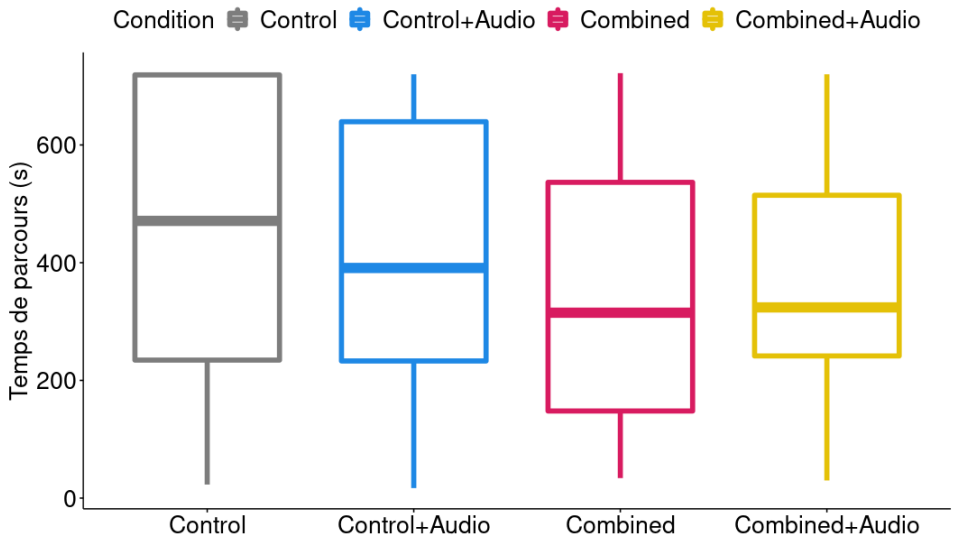


FIGURE 3.15 – Temps de parcours par condition. Les résultats ne sont pas significatifs. (N=23. Les barres indiquent l'intervalle de confiance à 95%. *=0.05 ; **=0.01 ;**=0.001 ; ****=0.00001)

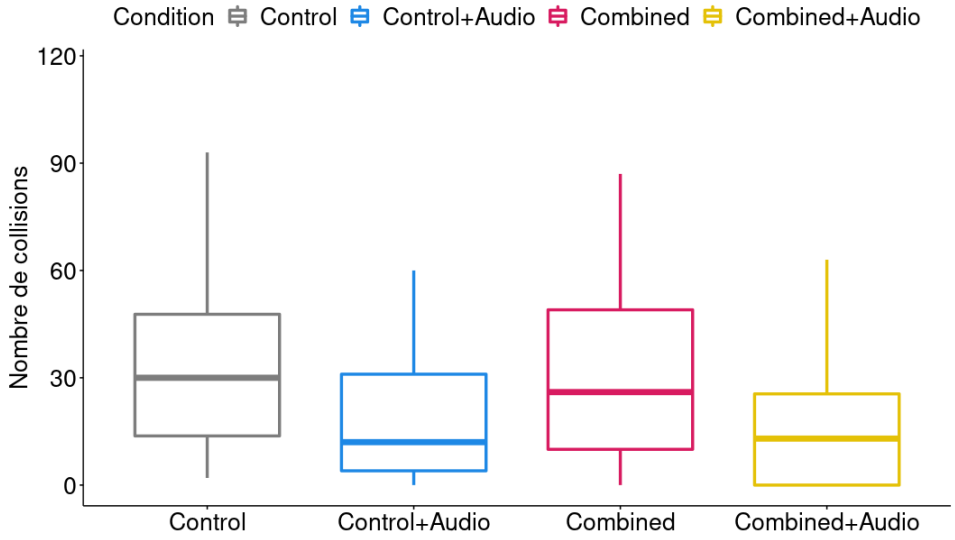


FIGURE 3.16 – Nombre de collisions par condition. Les résultats ne sont pas significatifs. (N=23. Les barres indiquent l'intervalle de confiance à 95%. *=0.05 ; **=0.01 ;**=0.001 ; ****=0.00001)

$p < 0.00001$). Un test post-hoc de Tukey montre les comparaisons par paires dans la FIGURE 3.17.

Conformément à notre hypothèse, on constate que lorsque le feedback audio est fourni, le nombre d'objets correctement reconnus est significativement plus élevé.

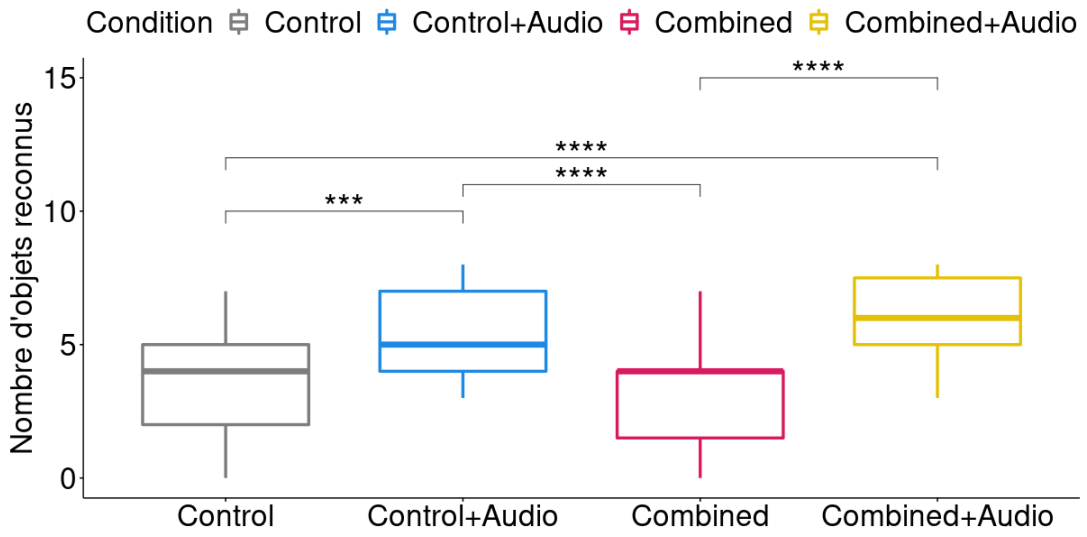


FIGURE 3.17 – Les résultats sont significativement différents sur le nombre d'objets bien reconnus selon les conditions. (N=23. Les barres indiquent l'intervalle de confiance à 95%. *=0.05 ; **=0.01 ; ***=0.001 ; ****=0.00001).

Localisation des objets

La localisation des objets est évaluée par le nombre d'objets reconnus bien localisés sur la carte de la scène après la navigation. Une ANOVA à un seul facteur (Condition) a montré que le nombre d'objets bien localisés est significativement différent selon les conditions ($F(3, 84) = 13.093$; $p < 0.00001$). Un test post-hoc de Tukey montre que les comparaisons par paires sur la FIGURE 3.18.

Nous pouvons voir ici que les objets sont significativement mieux localisés lorsque des indices sonores sont présents. De plus, les objets sont également significativement mieux placés lorsque l'on utilise la condition de rendu *Combiné*+Audio par rapport à la condition Contrôle+Audio.

Identification de la scène

L'identification de la scène est évaluée en vérifiant si la scène choisie par le sujet est la bonne. Les sujets devaient choisir une configuration parmi 4. Une ANOVA à sens unique (Condition) a montré que l'identification de la scène n'est pas significativement différente entre les conditions ($F(3,84) = 0.75$; $p = 0.55$). On peut noter que pour les 4 conditions, la moyenne de bonne reconnaissance est supérieure au seuil de chance (Contrôle = 47%, Contrôle + Audio = 70%, *Combiné* = 40% et *Combiné* + Audio = 56%).

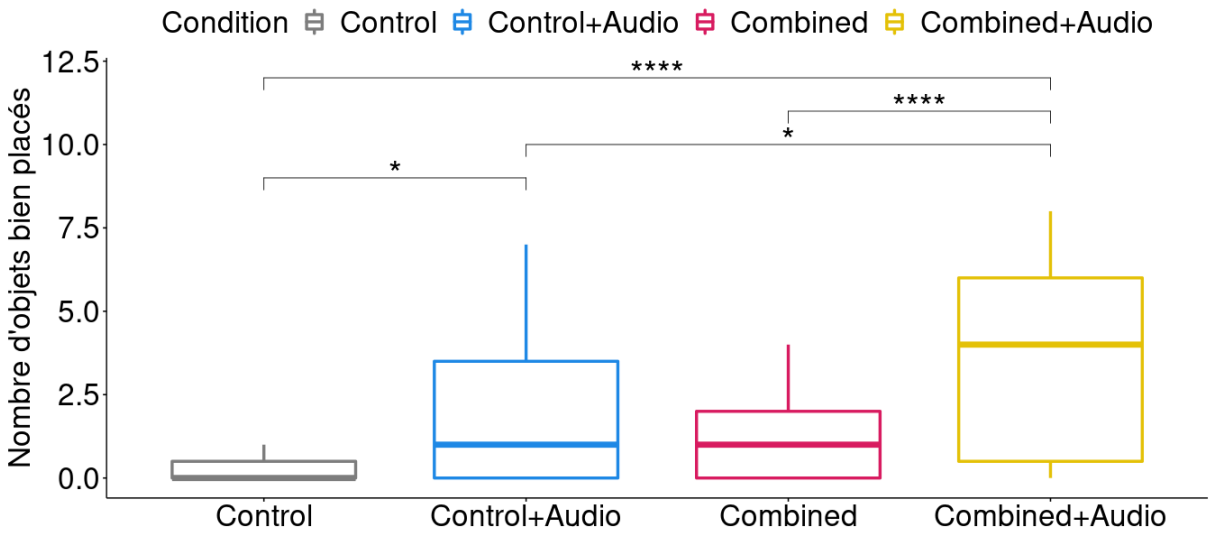


FIGURE 3.18 – Les résultats sont significativement différents sur le nombre d'objets bien localisés selon les conditions. (N=23. Les barres indiquent l'intervalle de confiance à 95%. *=0.05 ; **=0.01 ; ***=0.001 ; ****=0.00001).

3.4.3 Hypothèse 3 : Préférences de l'utilisateur

MeCue

Pour évaluer la satisfaction des sujets pour les 4 conditions, on leur soumet un formulaire MeCue. Un score de satisfaction est alors calculé à partir des réponses fournies par les sujets. Ce score est réparti sur une plage de valeurs allant de 0 à 100 (0 étant très insatisfait et 100 très satisfait). Une ANOVA à un seul facteur (Condition) sur l'indice de satisfaction montre que les résultats sont significativement différents selon la condition ($F(3,88) = 5.45$, $p = 0.002$) comme le montre la FIGURE 3.19.

Les résultats montrent que les rendus avec des indices audio sont préférés au rendu de contrôle. En revanche, ils ne montrent pas de différences nettes entre les autres paires de conditions.

Classement final

Pour évaluer les classements, nous procédons comme suit : un rendu classé en première position rapporte 4 points, un rendu classé en deuxième position rapporte 3 points, un rendu classé en troisième position rapporte 2 points et un rendu classé en quatrième position rapporte 1 point. Une ANOVA à un seul facteur (Condition) sur les classements montre que les résultats sont significativement différents selon les conditions ($F(3, 88) = 31{,}47$; $p < 0{,}0001$) comme le montre la FIGURE 3.20.

Les résultats montrent que les sujets ont évalué les rendus avec des repères audio significativement mieux que ceux sans repères. Nous notons également qu'il n'y a pas de

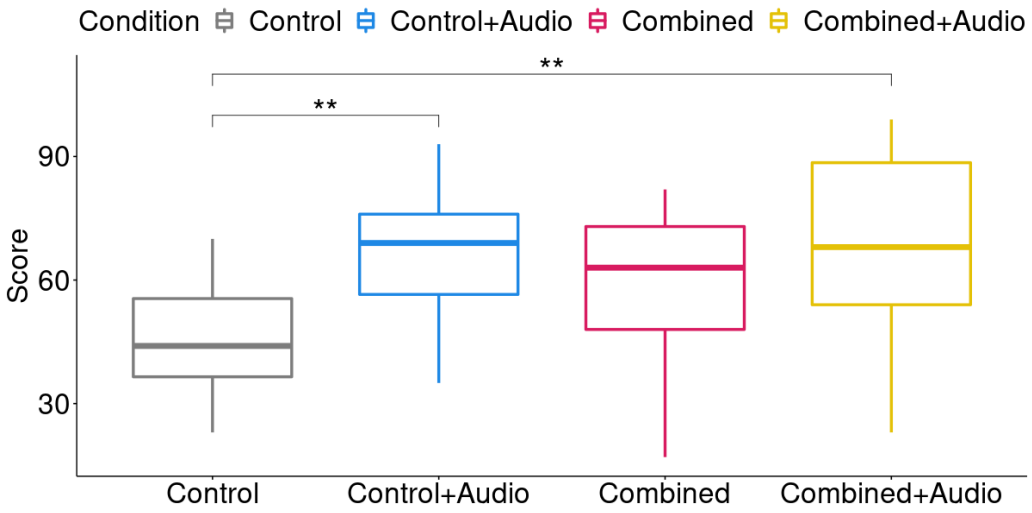


FIGURE 3.19 – Les résultats sont significativement différents sur l'indice de satisfaction selon les conditions. (N=23. Les barres indiquent l'intervalle de confiance à 95%. *=0.05 ; **=0.01 ; ***=0.001 ; ****=0.00001).

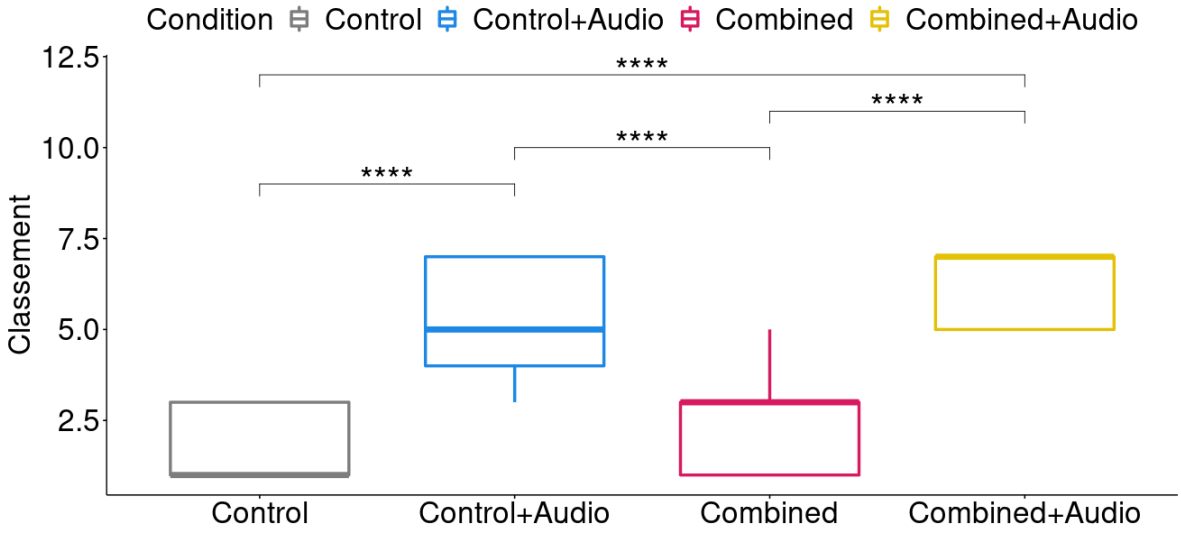


FIGURE 3.20 – Les résultats sont significativement différents sur le classement selon les conditions. (N=23. Les barres indiquent un intervalle de confiance de 95%. *=0.05 ; **=0.01 ; ***=0.001 ; ****=0.00001).

différence significative entre Contrôle et *Combiné* ou entre *Scoreboard*+Audio et *Combiné*+Audio.

3.5 Analyse des résultats et discussion

3.5.1 Comportement des sujets

On remarque tout d'abord que ni le rendu, ni l'audio n'ont un impact significatif sur les collisions avec les objets et les murs, ou sur le temps de navigation. En ce qui concerne le temps de navigation, cela peut s'expliquer par le fait que dans le cas d'un rendu jugé difficile par un utilisateur, un temps de navigation long peut signifier une réelle difficulté d'utilisation. Mais comme la tâche consiste à localiser des objets et à atteindre la sortie le plus rapidement possible, dans le cas d'un rendu jugé plus facile, les sujets passent plus de temps à chercher des objets, même si la sortie a déjà été trouvée.

3.5.2 Capacités d'analyse et de perception du sujet

D'après les résultats ci-dessus, il est clair que l'audio a un impact significatif sur le nombre d'objets reconnus et localisés. Le rendu ne semble pas avoir d'impact significatif sur les performances de reconnaissance et de localisation des objets. Il semble que ce soit la conjonction de perceptions visuelles et auditives qui permette d'obtenir les meilleurs résultats.

Concernant l'identification de la scène, les résultats ne sont pas non plus significatifs. Cette tâche étant la plus difficile puisque c'est celle qui demande le plus d'investissement cognitif (reconnaissance des joints de la pièce, mur-plafond, sol-mur), il est possible que les sujets ne se soient pas concentrés sur cette tâche, mais plus sur le fait de trouver, reconnaître et localiser des objets dans la scène.

3.5.3 Satisfaction et classement

Nous pouvons clairement voir que la préférence des sujets va vers les rendus avec audio, ce qui se reflète dans les classements finaux ainsi que dans les commentaires des sujets à la fin de l'expérience. Les sujets décrivent un bien meilleur confort de navigation en utilisant les rendus combinés à l'audio qu'en utilisant les rendus seuls. Parmi les sujets ayant laissé des commentaires, certains soulignent que l'audio les a aidés à comprendre la composition des différentes pièces qui composent la scène, ce qui leur a permis de s'orienter et donc de choisir la prochaine pièce dans laquelle ils souhaitaient aller avec plus de confiance.

3.5.4 Concernant les différents type de rendus

Dans cette étude, nous montrons l'importance de l'audio sur la tâche de navigation. Mais qu'en est-il de l'importance du rendu ? Dans la tâche que nous avons construite, il semble que le rendu seul ne soit pas un facteur suffisant pour assurer une bonne détection et

localisation des objets. En revanche, il n'est pas inutile pour autant. Il permet, comme l'ont souligné plusieurs sujets dans les commentaires, d'essayer de construire une représentation mentale de la scène.

Nous proposons deux explications à ces observations. La première est que l'apport de l'audio pour cet exercice est tel qu'il éclipse complètement l'intérêt d'un rendu sophistiqué. La seconde est que la condition *Combiné*+Audio n'a jamais été testée en simulation et qu'il est en fait difficile d'appréhender un tel rendu, aussi informatif soit-il. On pourrait imaginer qu'avec des expériences répétées, il serait possible de mesurer un effet d'apprentissage sur différents rendus.

3.5.5 Concernant la simulation

Notre environnement est très simple (encombrement faible, grands espaces entre chaque objet). Nous avons montré que les techniques de rendu utilisées dans le cadre de cette étude présentent un intérêt négligeable pour la compréhension de la scène par rapport à l'ajout d'un retour audio, mais nous pourrions maintenant rendre notre scène plus complexe pour voir si cette proposition reste vraie même dans un environnement plus riche. Nous pourrions également réaliser une étude similaire dans un environnement extérieur avec des obstacles mobiles.

Un moyen de tester et valider cette expérience dans des conditions plus proches de la réalité serait de mettre en place le système suivant : un casque de réalité virtuelle équipé d'une caméra qui réalise l'acquisition de la scène et un boîtier de calcul en temps réel (smartphone par exemple). Le flux vidéo est ensuite transmis à l'unité de calcul, sur laquelle deux réseaux de neurones sont embarqués pour prédire respectivement les cartes de segmentation sémantique et les cartes de structure qui sont ensuite utilisées pour générer le rendu final affiché sur les écrans du casque. Un environnement réel de simulation composé de divers éléments tels que des tables, des chaises ou des bureaux doit ensuite être construit. Cependant, il serait probablement nécessaire d'évaluer le pourcentage d'erreurs dans la prédiction des cartes de segmentation sémantique, que ce soit en termes de pixels ou d'étiquettes associées. Pour cela, nous avons besoin de réseaux entraînés qui soient suffisamment robustes pour limiter les erreurs. Ceci dit, le plus important est d'éviter les erreurs dans la prédiction des étiquettes qui peuvent être utilisées pour générer les signaux audio.

3.5.6 Concernant les retours audio

Dans le cadre de cette étude, les indices sonores possibles ont été préalablement listés et enregistrés par un logiciel de synthèse vocale. Étant donné que l'étude se déroule en simulation contrôlée, ce choix semblait judicieux. En revanche, pour une mise en œuvre réelle, il serait nécessaire de réfléchir à la manière dont on pourrait intégrer ce dispositif en

conditions réelles, face à des objets de catégories plus variées. Il nous semble que la bonne façon de procéder est de s'appuyer sur les prédictions faites par les réseaux de neurones embarqués et d'utiliser les étiquettes des objets reconnus comme source pour les indices audio.

En résumé

Dans ce chapitre, nous avons étudié l'intérêt d'ajouter des indices audio couplés à une restitution visuelle dans un cadre de navigation en intérieur. Cette étude se basait principalement sur l'analyse des besoins des non-voyants établie par Ratelle et Couturier [Ratelle, 2019].

Nous avons dans un premier temps présenté plusieurs dispositifs trouvés dans la littérature utilisant des systèmes de communication d'informations pour les non-voyants basés sur les retours audio. Dans un deuxième temps, nous avons détaillé notre protocole expérimental. Enfin, nous avons mis en évidence des résultats significatifs qui permettent de répondre à notre hypothèse initiale, qui était pour rappel que l'ajout d'indications sonores doit permettre aux sujets en situation de navigation de mieux comprendre et percevoir leur environnement.

D'après les résultats présentés, le fait d'ajouter des indices audio dans des tâches de navigation en intérieur permet de désambiguïser la scène et de fournir une aide significative aux sujets en simulation. Cela ne signifie pas pour autant que le système de rendu prothétique visuel ne sert à rien.

Dans le prochain chapitre, nous aborderons un aspect différent de la restitution visuelle par neuroprothèse. Nous cherchons à savoir si l'ajout d'un système de sélection de rendus en temps réel permet d'améliorer la compréhension de la scène. En d'autres termes, nous cherchons à savoir si une neuroprothèse interactive permet aux sujets de mieux comprendre leur environnement. Cette étude est basée sur le fait que nous avons obtenu des résultats mitigés sur la supériorité du rendu *Combiné* vis-à-vis du rendu *Scoreboard*. Nous avons souhaité étudier plus en détail sa complexité en donnant la possibilité aux utilisateurs de le simplifier interactivement.

Chapitre 4

Une neuroprothèse visuelle interactive pour mieux percevoir l'environnement

Sommaire

4.1 Matériel et méthodes . . . 109
4.1.1 Simulateur de vision prothétique . . . 109
4.1.2 Rendus phosphéniques . . . 114
4.1.3 Hypothèse . . . 114
4.1.4 Sujets . . . 115
4.1.5 Conditions expérimentales . . . 115
4.2 Protocole . . . 116
4.2.1 Protocole expérimental . . . 116
4.2.2 Variables analysées . . . 117
4.3 Résultats . . . 119
4.3.1 Comparaison des trois rendus utilisés . . . 119
4.3.2 Jugement subjectif . . . 123
4.3.3 Classement final . . . 124
4.3.4 Analyse détaillé du comportement observé pour le rendu Switch . . . 124
4.4 Analyse et discussion des résultats . . . 126
4.5 Limitations de l'étude et perspectives . . . 129

Dans le chapitre précédent (chapitre 3), nous nous sommes intéressés à l'ajout d'indices sonores dans un contexte de navigation intérieure. Cet ajout a permis de montrer la supériorité de l'audio par rapport à l'utilisation d'un rendu sophistiqué (combinant la sémantique et la structure). On peut alors se demander dans quelle mesure des rendus conçus pour répondre à certaines tâches peuvent être quand même utilisés. Cette interrogation a donné naissance à l'idée selon laquelle laisser le choix à l'utilisateur du rendu utilisé pourrait lui permettre d'améliorer ses performances lors d'une navigation.

De nombreuses études montrent que des rendus spécifiques peuvent améliorer la perception d'éléments précis de la scène [Vergnieux et al., 2014, Vergnieux et al., 2017, Li et al., 2018]. Cependant, une scène visuelle n'est pas constituée d'un seul type d'éléments. Elle comprend généralement une structure et des objets répartis dans la scène. Il est donc logique de penser que la combinaison de plusieurs types de rendus peut améliorer la perception de la scène. En effet, Sanchez-Garcia et al. [Sanchez-Garcia et al., 2020] ont proposé un rendu qui construit une représentation schématique des environnements intérieurs, mettant en évidence les bords informatifs structurels et les silhouettes des objets segmentés. Un exemple d'une telle restitution est visible sur la FIGURE 4.1. Ils ont montré à l'aide d'un simulateur de vision prothétique (32x32 électrodes, 8 niveaux de luminance), que ce rendu est supérieur au rendu scoreboard pour la reconnaissance des objets et l'identification des pièces.

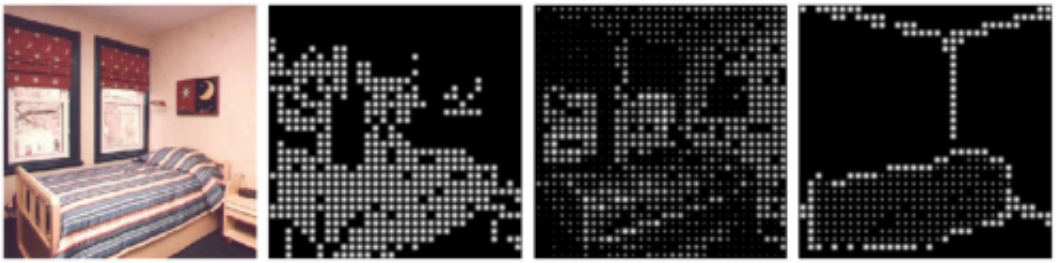


FIGURE 4.1 – Exemple de rendu basé sur l'approche proposée par Sanchez-Garcia et al. Les images représentent dans l'ordre, l'image de base, une restitution phosphénique basée uniquement sur une détection par filtre de Sobel, une restitution phosphénique basé sur une technique du type *Scoreboard* et l'image proposée par leur méthode SIE-OMS [Sanchez-Garcia et al., 2020].

Notre précédente étude (chapitre 3) a mis en exergue que le rendu *Combiné* n'est pas meilleur que le rendu *Scoreboard*, indépendamment de la présence des indices audio, ce qui contredit les résultats de Sanchez-Garcia. Notre avis est que les résultats obtenus par Sanchez-Garcia ne sont valides qu'à haute résolution. Dès que la résolution diminue, ce résultat peut devenir invalide. Une solution pour conserver les bénéfices du rendu combiné à faible résolution (objets et structure mis en évidence) consisterait à donner à l'utilisateur la possibilité de passer d'un rendu à l'autre selon son désir (rendu interactif). Dans ce cas, l'encombrement visuel serait réduit et la perception améliorée.

Nous avons construit un protocole expérimental ainsi qu'une tâche visuelle consistant à identifier des objets et des configurations de carrefours routiers dans des scènes visuelles virtuelles. Cette étude a pour but de montrer l'intérêt de fournir aux non-voyants implantés un outil de sélection de rendus en temps réel pour les aider à mieux comprendre et percevoir leur environnement. Dans le cadre de cette étude, nous avons choisi de modéliser des environnements extérieurs pour compléter l'étude précédente qui traitait déjà des environnements intérieurs. De cette manière nous proposons une seconde étude qui met en évidence les besoins des non-voyants exprimés dans le rapport de Ratelle et Couturier [Ratelle, 2019] en situation de navigation extérieure. Les réponses de vingt sujets ont confirmé notre hypothèse selon laquelle le rendu interactif améliore significativement la compréhension de la scène. La suite s'articule de la manière suivante : une présentation du matériel et des méthodes utilisées (4.1), la description du protocole expérimental (4.2), la présentation des résultats (4.3), des commentaires sur les résultats (4.4) et enfin nous présenterons quelques limitations (4.5).

4.1 Matériel et méthodes

Nous avons utilisé le simulateur que nous avons présenté au chapitre 2 pour réaliser cette expérience. La situation sanitaire exceptionnelle de ces dernières années nous a conduit à devoir réaliser cette expérience à distance. C'est pourquoi nous avons dû construire une version Web de notre simulateur.

4.1.1 Simulateur de vision prothétique

Gestion du module Web

Dans le cadre de cette étude, le simulateur a été compilé en une version WebGL de manière à être rendu disponible sur Internet. Les sujets réalisant l'étude doivent uniquement interagir avec le simulateur grâce au clavier et à la souris. Les réponses aux questions sont récupérées grâce à l'envoi d'une requête POST sur un serveur Apache déployé sur une Raspberry pi 4. Une fois les données reçues, elles sont enregistrées dans une base de données liée au serveur Apache.

Modèle 3D utilisé

Nous avons utilisé le pack d'asset 3D, permettant de créer des environnements extérieurs, proposé par Paul PJ Maget (City Voxel Pack [1]). Il s'agit d'un ensemble de bâtiments et de rues interconnectés. Un exemple de scène utilisée dans notre étude est montré sur la Figure 4.2.

1. https://paulpjmaget.wixsite.com/poeme

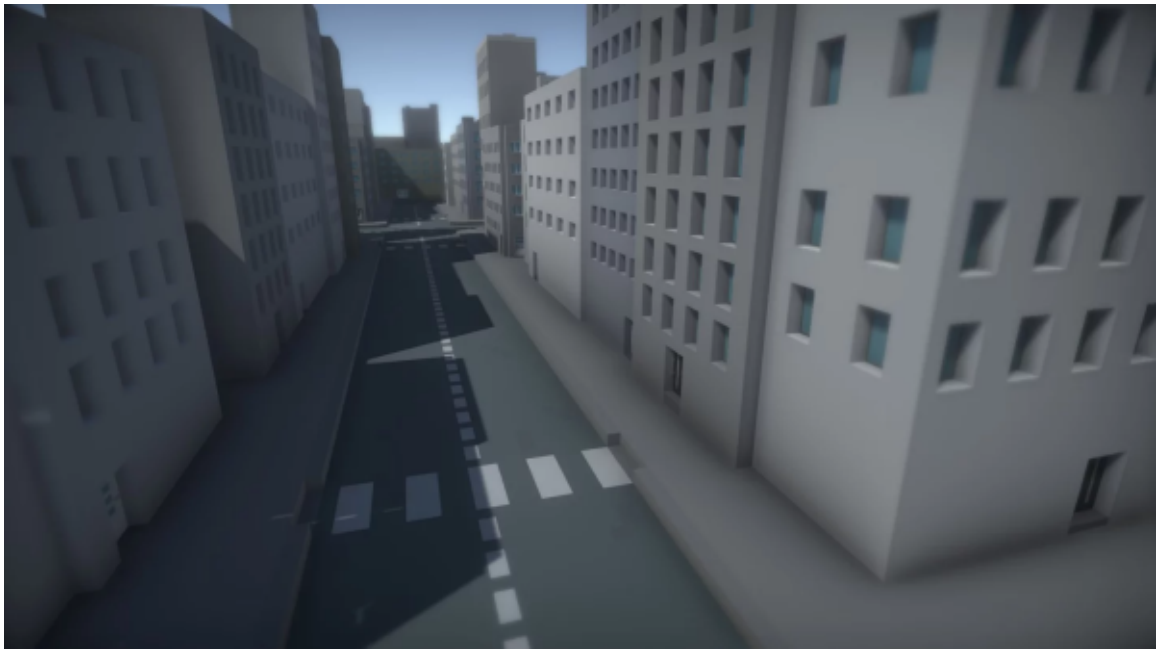


FIGURE 4.2 – Exemple de scène utilisée lors de l'étude. On retrouve ici une longue avenue traversée par un passage piéton. Des immeubles bordent la route.

Nous avons également placé quelques obstacles pour rendre la scène plus réaliste (voitures, bancs, arbres, passages piétons, portes). Le choix de ces obstacles est motivé par l'analyse du rapport de Ratelle et Couturier [Ratelle, 2019] qui montre que les besoins les plus importants pour les déficients visuels en situation extérieure sont la détection d'obstacles (voitures, bancs, arbres), la détection de coins de rues et d'intersections (bâtiments) et la détection d'entrées d'immeubles (portes) et de passages piétons. Comme nous l'avons déjà mentionné dans les chapitres précédents, les besoins émis par les non-voyants se déclinent en deux catégories, les besoins intérieurs et les besoins extérieurs. La simulation en extérieur permet de facilement manipuler des éléments dynamiques (par exemple des voitures) alors que dans des scènes d'intérieur, il est rare d'avoir des éléments mobiles.

Construction des questions

Dans cette étude, nous allons présenter une série de questions aux sujets pour évaluer leur compréhension visuelle de la scène. Ces questions sont proposées sous différentes restitutions visuelles. La résolution choisie pour la modélisation de l'implant est de 21x18 est sera justifiée en 4.1.1. Les tableaux 4.1, 4.2 et 4.3 montrent les images utilisées dans l'expérience, respectivement dans les catégories de questions objets, rues, et portes et passages piétons.

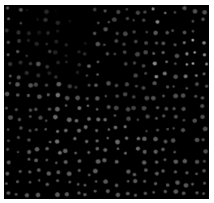
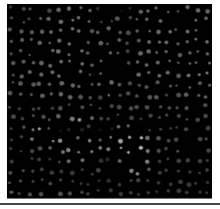
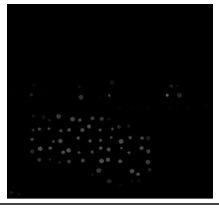
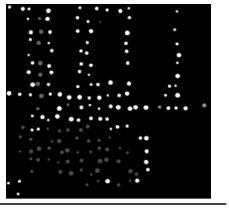
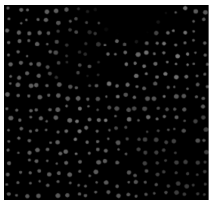
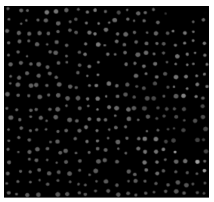
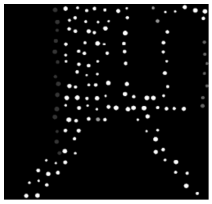
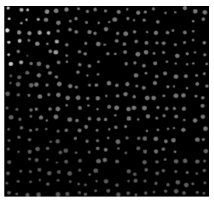
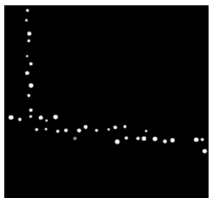
(a) Scène 3D	(b) Scoreboard	(c) Segmentation	(d) Structure	(e) Combiné
				
				
				
				
				

TABLEAU 4.1 – Exemples de rendus prothétiques : (a) montre l'image initiale dans un environnement virtuel. (b) montre comment (a) est rendu avec le rendu *Scoreboard*. (c) montre comment (a) est rendu avec une méthode qui détecte uniquement les objets (segmentation sémantique des objets), et (d) montre comment (a) est rendu avec une méthode qui améliore les informations structurelles (amélioration de la structure). (e) est un rendu combinant (c) et (d) et appelé *Combiné*. La résolution de l'implant simulée est de 21x18.

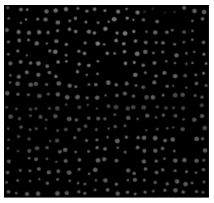
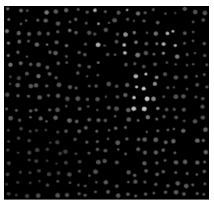
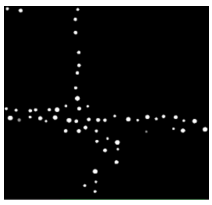
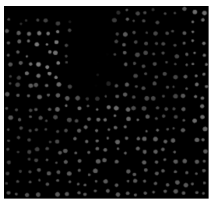
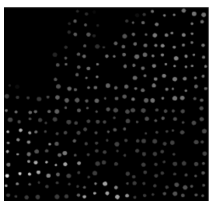
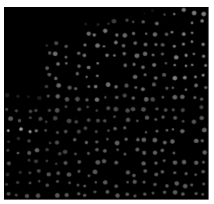
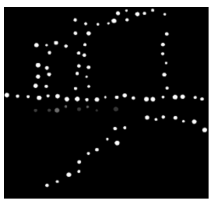
(a) Scène 3D	(b) Scoreboard	(c) Segmentation	(d) Structure	(e) Combiné
				
				
				
				
				

TABLEAU 4.2 – Exemples de rendus prothétiques : (a) montre l'image initiale dans un environnement virtuel. (b) montre comment (a) est rendu avec le rendu *Scoreboard*. (c) montre comment (a) est rendu avec une méthode qui détecte uniquement les objets (segmentation sémantique des objets), et (d) montre comment (a) est rendu avec une méthode qui améliore les informations structurelles (amélioration de la structure). (e) est un rendu combinant (c) et (d) et appelé *Combiné*. La résolution de l'implant simulée est de 21x18.

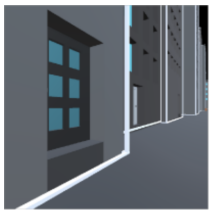
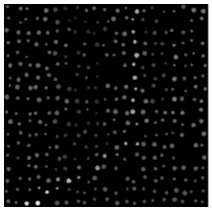
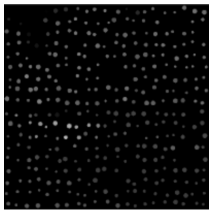
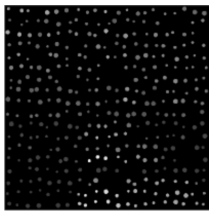
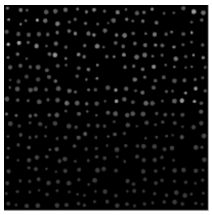
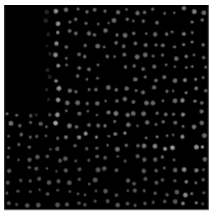
(a) Scène 3D	(b) Scoreboard	(c) Segmentation	(d) Structure	(e) Combiné
				
				
				
				
				

TABLEAU 4.3 – Exemples de rendus prothétiques : (a) montre l'image initiale dans un environnement virtuel. (b) montre comment (a) est rendu avec le rendu *Scoreboard*. (c) montre comment (a) est rendu avec une méthode qui détecte uniquement les objets (segmentation sémantique des objets), et (d) montre comment (a) est rendu avec une méthode qui améliore les informations structurelles (amélioration de la structure). (e) est un rendu combinant (c) et (d) et appelé *Combiné*. La résolution de l'implant simulée est de 21x18.

Implant simulé

Nous avons choisi de simuler le système PRIMA développé par Pixium Vision (Paris, France), composé d'un réseau de 21x18 électrodes. Notre choix est motivé par le fait que c'est l'implant le plus récent ayant passé des phases d'essais cliniques satisfaisantes [Fernandez, 2018]. Il est aussi l'un de ceux qui permet un rendu visuel avec la plus grande résolution à ce jour. En effet, les limitations techniques actuelles ne permettent pas d'avoir un nombre d'électrodes très élevé (l'Argus II possède seulement 6x10 électrodes [Farvardin et al., 2018]). Ce nombre est largement inférieur à celui des implants simulés dans de nombreuses études de la littérature [Li et al., 2018, Sanchez-Garcia et al., 2020], tableau 1.4.3. Pendant l'expérience, pour s'assurer que le rendu visuel était le plus réaliste possible, nous avons demandé aux sujets de se placer à une distance comprise entre 50 et 75 *cm* de l'écran. Étant donné que la zone d'affichage sur l'écran était de 10 *cm* x 10 *cm*, à une telle distance le champ de vision perçu est de l'ordre de 10°, ce qui est proche de celui mesuré dans les essais cliniques du système PRIMA. Tout comme dans la première étude (chapitre 3), un taux de défaillance (appelé dropout en anglais) de 10% était appliqué aux électrodes, afin de simuler des électrodes non-fonctionnelles ou cassées. Un flou gaussien était appliqué aux phosphènes générés, la taille des phosphènes au sein d'un même implant variait entre 0,235° et 0,275° du champ de vision. L'espacement entre deux phosphènes variait entre 0,55° et 0,825° du champ de vision. Chaque électrode pouvait générer un phosphène avec quatre niveaux de gris.

4.1.2 Rendus phosphéniques

Dans cette étude, nous avons utilisé le rendu *Scoreboard* (Contrôle), utilisé par défaut dans certains implants commercialisés actuellement. Nous avons aussi utilisé un rendu mélangeant les segmentations d'objets et de structure (appelé *Combiné*) récemment proposé par Sanchez et al. [Sanchez-Garcia et al., 2020]. Pour tester notre hypothèse, nous avons conçu un troisième mode de rendu appelé *Switch* qui permet de changer le rendu à volonté parmi les trois suivants : objets seuls, structure seule, objets + structure. Nous avons détaillé la construction des rendus *Scoreboard* et *Combiné* dans le chapitre 2. La nouveauté est ici le mode de rendu *Switch* qui permet au sujet de changer de rendu à volonté. Il lui suffit, pour ce faire, de cliquer sur un bouton prévu pour cet effet (voir FIGURE 4.3). Dans ce mode, les rendus disponibles sont le rendu *Combiné*, le rendu objets et le rendu structure.

4.1.3 Hypothèse

Nous avons formulé l'hypothèse suivante : la possibilité de changer de type de rendu en temps réel pour réaliser des tâches de compréhension de scènes visuelles extérieures permet aux sujets d'avoir de meilleures performances visuelles.

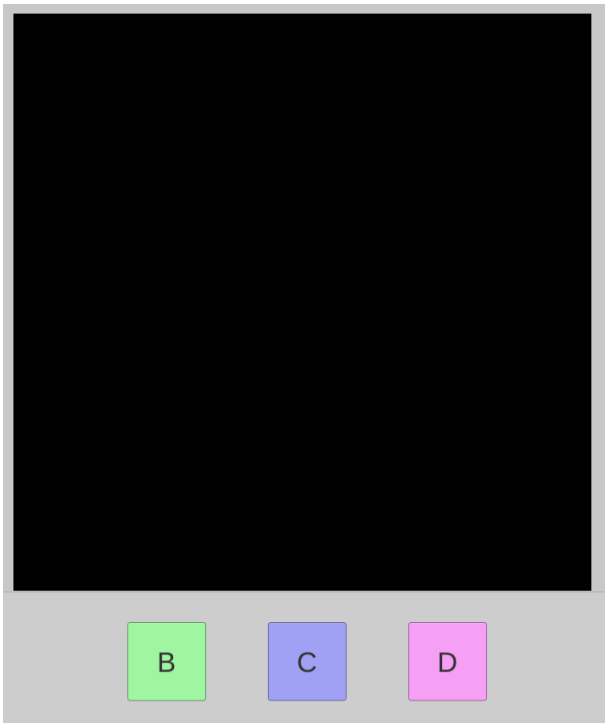


FIGURE 4.3 – Interface de sélection. Les boutons B, C et D permettent de changer dynamiquement de rendu. Dans ce cas précis, le bouton B permet d'activer le rendu d'objets seuls, le bouton C permet d'activer le rendu de structure seule et le bouton D permet d'activer le rendu *Combiné*.

4.1.4 Sujets

Les sujets ont été recrutés d'une part dans une école d'ingénieur et d'autre part via les réseaux sociaux (LinkedIn, Facebook). 20 sujets ont participé à l'expérience (11 hommes et 9 femmes âgés de 17 à 55 ans, moyenne : 25 ans, écart-type : 13 ans). Chaque participant avait une vision normale ou corrigée. Tous les sujets ont donné leur consentement pour participer à l'étude et permettre le stockage de leurs données anonymisées, en accord avec le RGPD.

4.1.5 Conditions expérimentales

Lors de cette, étude nous avons inclus trois conditions expérimentales :

— le rendu *Scoreboard* (Contrôle) seul ;
— le rendu *Combiné* (objets + structure) seul ;
— le mode *Switch* qui permet d'utiliser un ou plusieurs rendus (parmi objets, structure et *Combiné*) à volonté avant de répondre.

4.2 Protocole

4.2.1 Protocole expérimental

La session expérimentale était constituée de trois blocs correspondant aux trois conditions expérimentales. Chaque bloc était divisé en trois parties comprenant la familiarisation avec le rendu utilisé dans le bloc, la phase de test, puis un questionnaire concernant le rendu que le sujet vient d'utiliser dans le bloc.

Pendant la familiarisation, les sujets étaient libres d'utiliser le simulateur pour répondre à deux questions, sans contrainte de temps. Pour la condition *Switch*, ils pouvaient changer de rendu à volonté avant de répondre à ces deux questions.

Pendant la phase de test, le sujet devait répondre à une série de 15 questions incluant 5 questions dans les trois catégories nommées *objets*, *rues*, *portes et passages piétons*. Les questions de la catégorie *objets* portaient sur la présence d'objets qui peuvent gêner le déplacement du sujet. Les questions de la catégorie *rues* portaient sur la configuration des intersections. Une fois la réponse choisie, il passait à la question suivante. Les tableaux 4.4, 4.5 et 4.6 récapitulent les questions posées et les réponses attendues selon les catégories de questions.

Ces questions ont été choisies dans le but d'évaluer les critères de compréhension jugés les plus importants par Ratelle et Couturier [Ratelle, 2019]. On insère également volontairement des questions plus ambiguës pour éviter un biais de facilité.

L'ordre des blocs et des questions ont été triés de manière pseudo-aléatoire pour limiter les biais inter-blocs ou inter-questions. Pour éviter que les sujets ne reconnaissent certaines questions d'un bloc à un autre, l'angle de vue de la caméra dans la scène était modifié. Ainsi, la question restait sensiblement la même, mais la réponse pouvait changer selon l'orientation de la caméra. La FIGURE 4.4 montre l'interface de sélection de réponses.

À la fin de chaque bloc, les sujets devaient répondre à des questions subjectives (voir FIGURE 4.5) concernant la pertinence du rendu utilisé dans ce bloc. Les questions portaient sur le plaisir éprouvé et l'utilisabilité perçue concernant le rendu utilisé dans ce bloc. Il y avait 4 questions pour les conditions *Scoreboard* et *Combiné*, et 5 questions (dont 4 identiques aux précédentes) pour la condition *Switch*. Dans le cas de la condition *Switch*, on demandait également aux sujets s'ils trouvaient cette fonctionnalité pertinente. Pour toutes ces questions, ils devaient répondre sur une échelle de Likert allant de 1 à 7 (1 signifiant "pas du tout d'accord" et 7 "tout à fait d'accord"). Les questions posées se trouvent en annexe.

En fin d'expérience, les sujets devaient classer les différents rendus selon l'utilité perçue concernant les différentes tâches. La FIGURE 4.6 illustre l'interface de classement. L'étude est toujours disponible à l'adresse suivante : `http://ubee.enseeiht.fr/Julien/WebV2-17/index.html`. Les données ne sont cependant plus collectées.

Scène 3D	Question posée	Réponse
	Combien voyez-vous d'obstacles/objets ?	2
	A quel genre d'obstacle/objet pensez-vous faire face ?	Voiture
	A quel genre d'obstacle/objet pensez-vous faire face ?	Arbre
	Combien voyez-vous d'obstacles/objets ?	2
	Combien voyez-vous d'obstacles/objets ?	4

TABLEAU 4.4 – Questions posées et réponses attendues accompagnées des images de scènes réelles. Catégorie de questions "objets".

4.2.2 Variables analysées

Pour répondre à notre hypothèse, nous avons mesuré, pour chaque sujet, plusieurs variables :

- la validité de chaque réponse (correct / incorrect / je ne sais pas) ;
- le temps de réponse à chaque question ;
- pour la condition *Switch* :
  - l'ordre de sélection des rendus ;
  - la durée d'utilisation de chaque rendu ;
  - le nombre de changements de rendu.

Pour analyser les résultats quantitatifs nous avons effectué des tests ANOVA à 2 facteurs

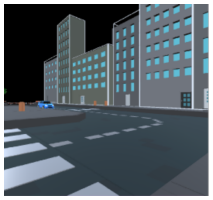
Scène 3D	Question posée	Réponse
	Dans quelle direction continue la route qui vous fait face ?	Vers la droite
	Dans quelle direction continue la route qui vous fait face ?	Vers la gauche
	Dans quelle direction continue la route qui vous fait face ?	Tout droit
	Vous trouvez- vous devant un coin de rue ?	Vers la gauche
	Vous trouvez- vous devant un coin de rue ?	Vers la droite

TABLEAU 4.5 – Questions posées et réponses attendues accompagnées des images de scènes réelles. Catégorie de questions "rues".

selon le modèle (Condition * Tâche * Interaction), puis nous avons utilisé un test post-hoc de Tukey pour comparer les paires deux à deux. Dans les figures, nous avons également affiché des intervalles de confiance à 95%. Le nombre d'étoiles sur les figures indique la confiance que l'on peut avoir dans les résultats. Ainsi, une étoile assure une confiance à 95%, deux étoiles une confiance à 99%, trois étoiles une confiance à 99,9% et quatre étoiles indiquent une confiance totale.

Scène 3D	Question posée	Réponse
	Vous trouvez-vous devant une porte ?	Oui
	Vous trouvez-vous devant une porte ?	Non
	Vous trouvez-vous devant un passage piéton ?	Oui
	Vous trouvez-vous devant un passage piéton ?	Non
	Vous trouvez-vous devant une porte ?	Non

TABLEAU 4.6 – Questions posées et réponses attendues accompagnées des images de scènes réelles. Catégorie de questions ”portes et passages piétons”.

4.3 Résultats

4.3.1 Comparaison des trois rendus utilisés

Réponses

Une ANOVA à deux facteurs (Condition * Tâche * Interaction) a montré que le nombre de réponses correctes est significativement différent selon le rendu ($F(2,171) = 15{,}834$; $p < 0{,}0001$) et selon la tâche $F(2,171) = 81{,}730$; $p < 0{,}0001$). L'interaction est également significative ($F(4,171) = 5{,}728$; $p < 0{,}001$), ce qui montre que la condition *Switch* est très efficace pour identifier les *objets*, mais encore plus pour comprendre la configuration des



FIGURE 4.4 – Exemple de question posée dans l'interface du simulateur. On retrouve ici les trois boutons de sélection de rendus, à savoir B pour le rendu d'objets, C pour le rendu de structure et D pour le rendu *Combiné*.

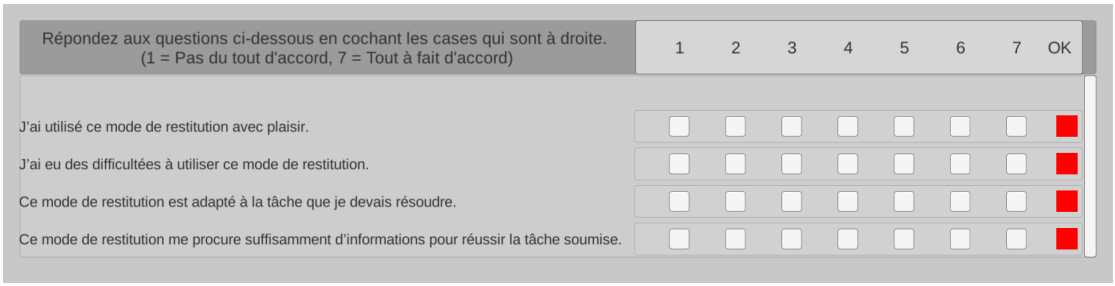


FIGURE 4.5 – Questions posées à l'issue de l'utilisation d'un rendu pendant l'étude.

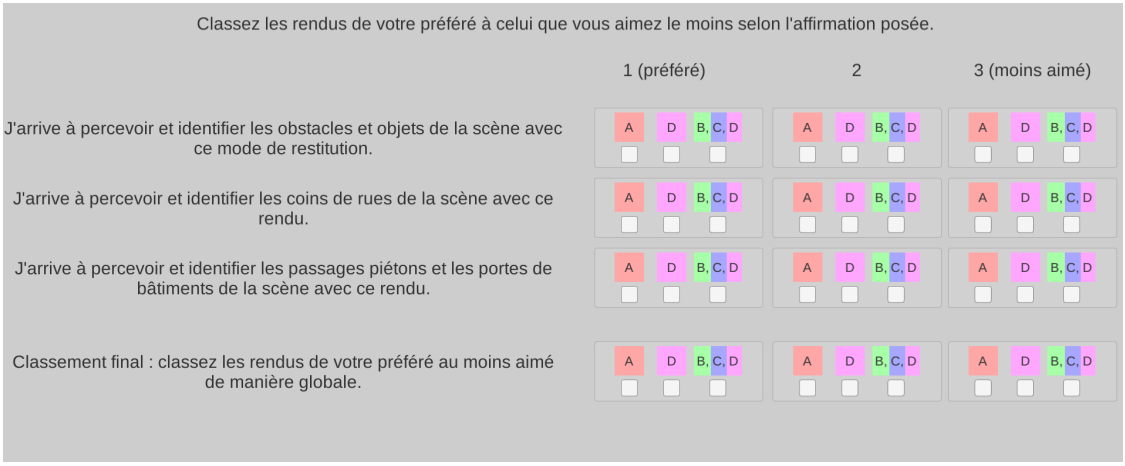


FIGURE 4.6 – Module de classement final.

rues. Un test post-hoc de Tukey montre que les résultats sont significatifs entre les trois rendus dans le cas des questions de type *rues* et entre les conditions *Switch* et Contrôle ou *Combiné*. La FIGURE 4.7 montre ces résultats.

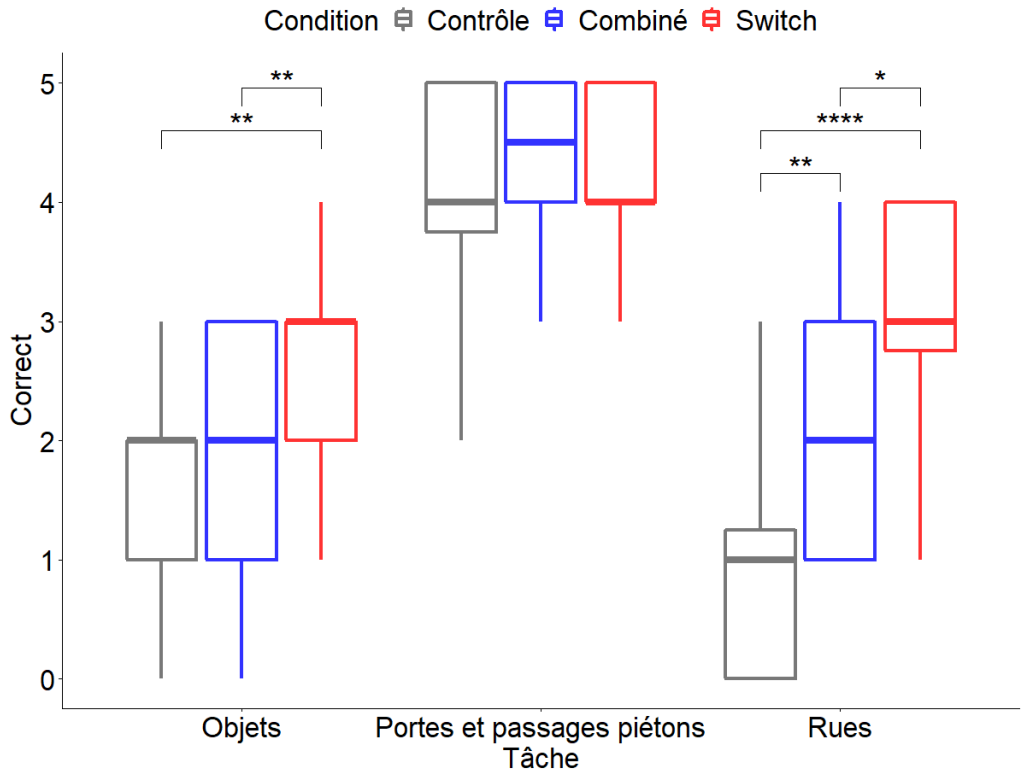


FIGURE 4.7 – Nombre moyen de réponses correctes par Rendu et par Tâche. La détection des passages piétons et des portes est équivalente pour toutes les conditions. L'identification des objets et des configurations de rues est significativement améliorée avec le rendu *Switch*. (N=20. Les barres indiquent un intervalle de confiance de confiance à 95%. *=0.05 ; **=0.01 ; ***=0.001 ;****=0.00001)

Réponses "Je ne sais pas"

Une ANOVA à deux facteurs (Condition * Tâche * Interaction) a montré que le nombre de réponses "je ne sais pas" est significativement différent selon le rendu ($F(2,171) = 10.543$; $p < 0{,}0001$) et selon la tâche $F(2,171) = 5.882$; $p < 0{,}001$). L'interaction n'est cependant pas significative ($F(4,171) = 1.909$; $p = 0.111$). Le test post-hoc de Tukey montre que les résultats sont significatifs entre les conditions Contrôle et *Switch* dans le cas de la tâche de type *objets* et entre les conditions Contrôle et *Combiné* ou *Switch* dans le cas de la tâche de type *rues*. La FIGURE 4.8 montre ces résultats.

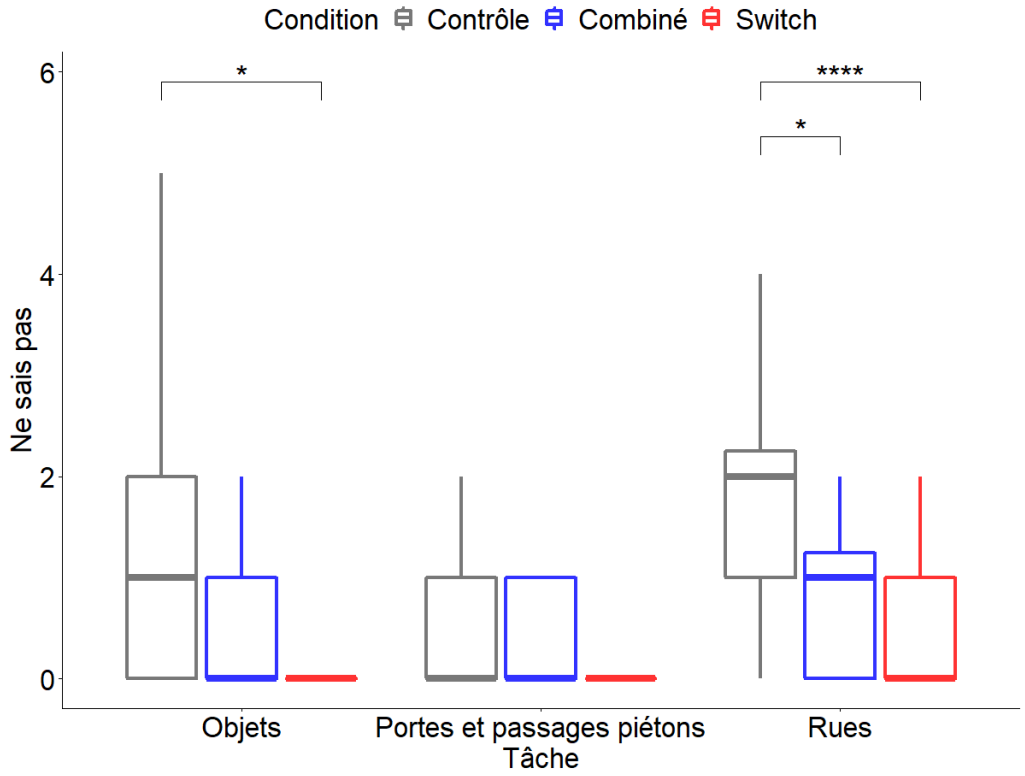


FIGURE 4.8 – Nombre moyen de réponses "je ne sais pas" par Rendu et par Tâche. Le nombre de réponses "je ne sais pas" est significativement plus élevé avec le rendu *Scoreboard* pour les tâches *objets* et *rues*. (N=20. Les barres indiquent un intervalle de confiance de confiance à 95%. *=0.05 ; **=0.01 ; ***=0.001 ; ****=0.00001).

Temps de réponse

La distribution des temps de réponse n'étant pas normale, nous avons réalisé une transformation logarithmique des données. Nous avons réalisé une ANOVA à deux facteurs (Condition * Tâche * Interaction) sur les données transformées qui a montré que le temps de réponse est significativement différent selon la condition expérimentale ($F(2,891) = 17.245$; $p < 0{,}0001$) et selon la tâche $F(2,891) = 14.081$; $p < 0{,}0001$). L'interaction n'est cependant pas significative ($F(4,891) = 0.465$; $p = 0.762$). Un test post-hoc de Tukey montre que les résultats sont significativement différents entre les conditions Contrôle et *Switch* dans le cas de la tâche de type *objets*, entre les conditions *Switch* et *Combiné* ou Contrôle dans le cas de la tâche de type *portes et passages piétons* et entre les conditions *Switch* et Contrôle dans le cas de la tâche *rues*. La FIGURE 4.9 montre ces résultats. Les sujets avaient un temps maximal pour répondre à chaque question fixé à 120 secondes. Au-delà de ce temps, la prochaine question était chargée. Notons que ce seuil était largement suffisant, étant donné qu'aucun sujet n'a dépassé un temps de réponse de 60 secondes.

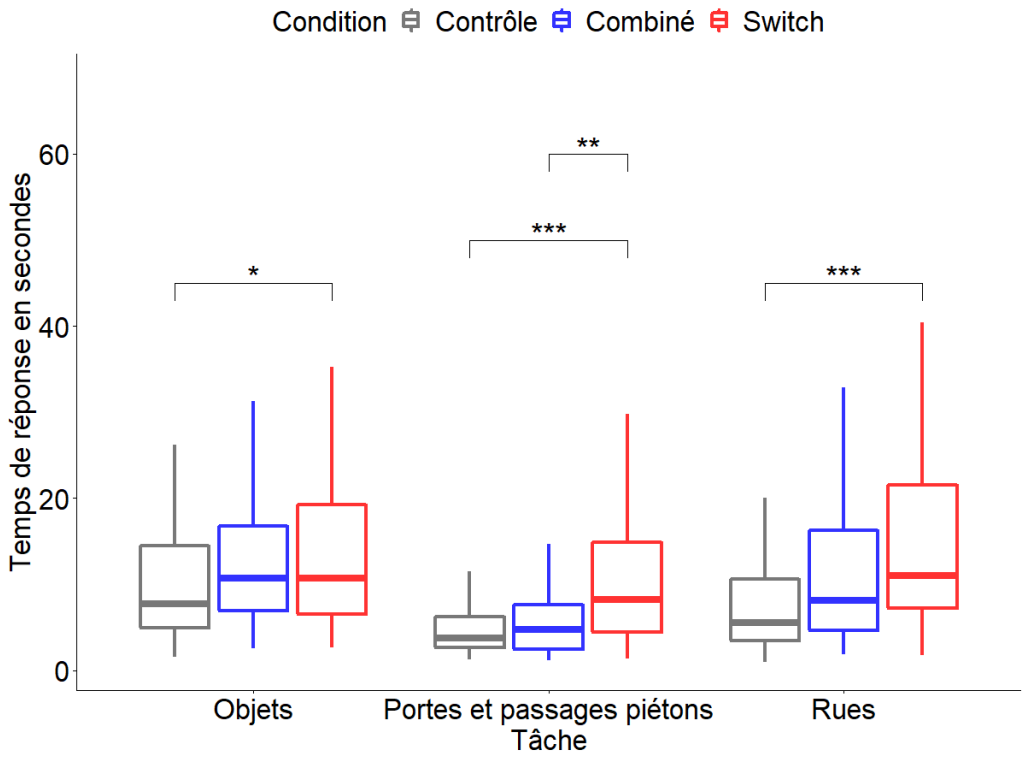


FIGURE 4.9 – Temps de réponse moyen par Rendu et par Tâche. Le rendu *Switch* obtient des temps de réponse plus longs comparés au rendu *Scoreboard*, quelle que soit la tache réalisée. (N=20. Les barres indiquent un intervalle de confiance à 95%. *=0.05 ; **=0.01 ; ***=0.001 ;****=0.00001).

4.3.2 Jugement subjectif

Pour rappel, les participants devaient répondre avec un score entre 1 et 7 à une série de questions à l'issue de chacun des trois blocs. Une ANOVA à deux facteurs (Rendu * Tâche * Interaction) montre que le score est significativement différent selon les rendus ($F(2,228) = 11.878$; $p < 0.0001$) et selon la tâche ($F(3,228) = 5.604$; $p < 0.001$). L'interaction est significative ($F(6,228) = 7.537$; $p < 0.0001$). Un test post-hoc de Tukey montre que les résultats sont significatifs pour les trois tâches.

La FIGURE 4.10 présente les résultats obtenus. On constate qu'en termes de plaisir d'utilisation, les rendus *Combiné* et *Switch* sont significativement mieux notés que le rendu de Contrôle (*Scoreboard*). On observe également que sur la difficulté d'utilisation, l'utilisabilité perçue et la quantité d'information présentée, le rendu *Switch* est mieux classé que le rendu de Contrôle. On remarque également que les rendus de Contrôle et *Combiné* ne sont pas perçus avec un niveau de difficulté différent. Notons qu'à la question : "J'ai trouvé la possibilité de changer de rendu très utile", on obtient un score moyen de 6.2 +/- 0.52 (sur 7).

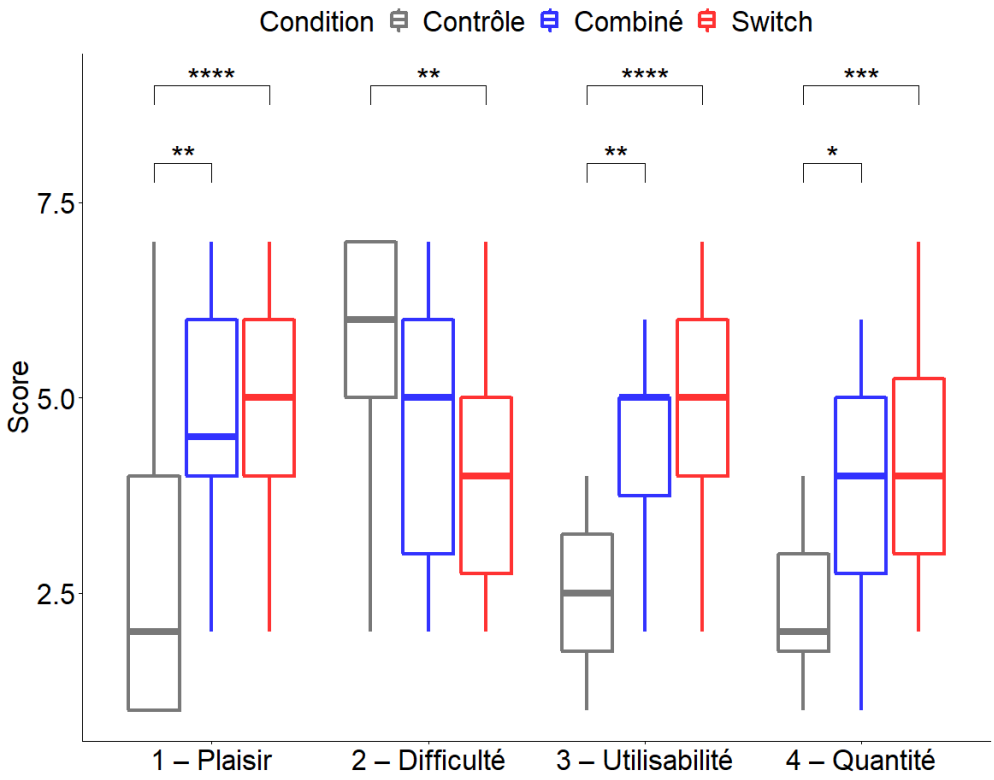


FIGURE 4.10 – Score moyen donné par les sujets aux trois rendus *Scoreboard*, *Combiné* et *Switch*. Cette figure montre des résultats significativement différents. Analyse réalisée sur les 20 sujets. Les scores sont compris entre 1 et 7 (1 signifiant pas du tout d'accord et 7 tout à fait d'accord).

4.3.3 Classement final

Les sujets devaient finalement classer les rendus par ordre de préférence en fonction de quatre dimensions : pour identifier les intersections et les coins de rues, pour identifier les objets, pour identifier les portes et passages piétons, et globalement.

Afin de générer un score de préférence, nous avons attribué 3 points si le rendu est classé premier, 2 points s'il est classé deuxième et 1 point s'il est classé troisième. La FIGURE 4.11 montre la moyenne de ces classements ainsi qu'une moyenne de ces quatre scores. On observe que l'ordre de préférence des rendus est *Switch*, *Combiné*, Contrôle, quelle que soit la tâche réalisée.

4.3.4 Analyse détaillé du comportement observé pour le rendu Switch

Dans le mode *Switch*, les utilisateurs pouvaient choisir d'utiliser librement un des trois rendus Objets, Structure, *Combiné*. Pour chaque essai réalisé par les sujets, nous avons analysé plus précisément le nombre de fois où chacun des rendus était utilisé, la durée

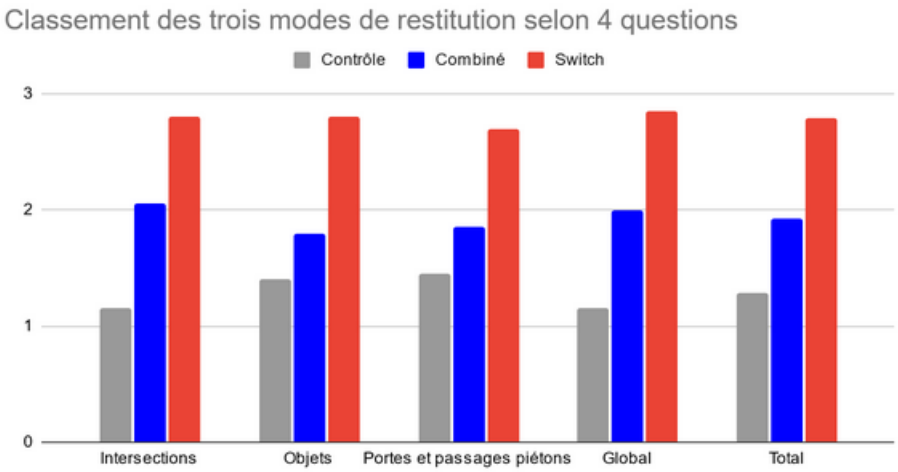


FIGURE 4.11 – Classement final moyen donné par les utilisateurs (N=20). Le rendu *Switch* est le mieux classé en moyenne, quelle que soit la tâche réalisée.

d'utilisation de chaque rendu et le dernier rendu utilisé.

La distribution des durées d'utilisation n'étant pas normale, nous avons réalisé une transformation logarithmique des données. Une ANOVA à un facteur (Tâche) a montré que les durées d'utilisation (après transformation) des rendus (Objets, Structure, *Combiné*) ne sont pas significativement différentes selon la tâche ($F(2,860) = 0.09$; $p = 0.914$). Ce résultat est montré sur la FIGURE 4.12.

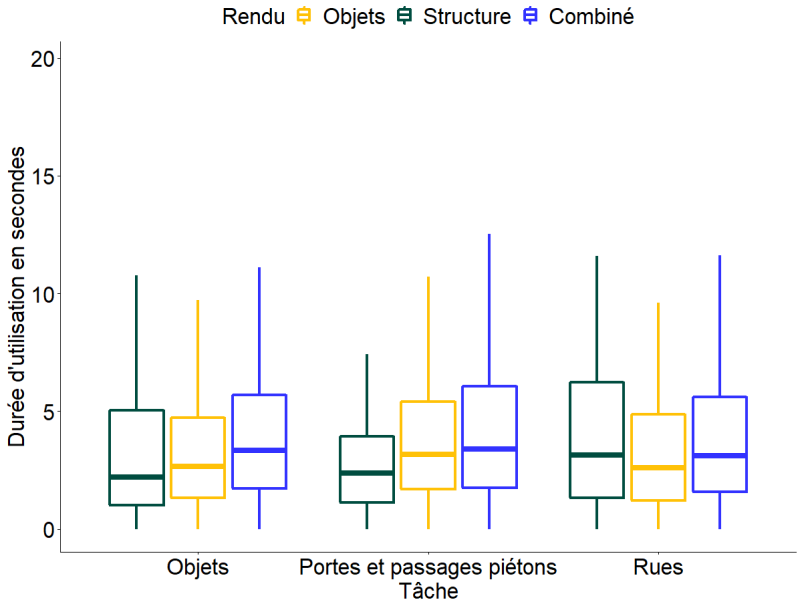


FIGURE 4.12 – Durée d'utilisation moyenne pour chacun des trois rendus du mode Switch par Tâche. Cette figure montre des résultats non significativement différents. Analyse réalisée sur les 20 sujets.

Une ANOVA à un facteur (Tâche) a montré que le nombre d'utilisations de chaque rendu (Objets, Structure, *Combiné*) n'est pas significativement différent selon la tâche non plus ($F(2,860) = 1.589$; $p = 0.205$). Ce résultat est montré sur la FIGURE 4.13.

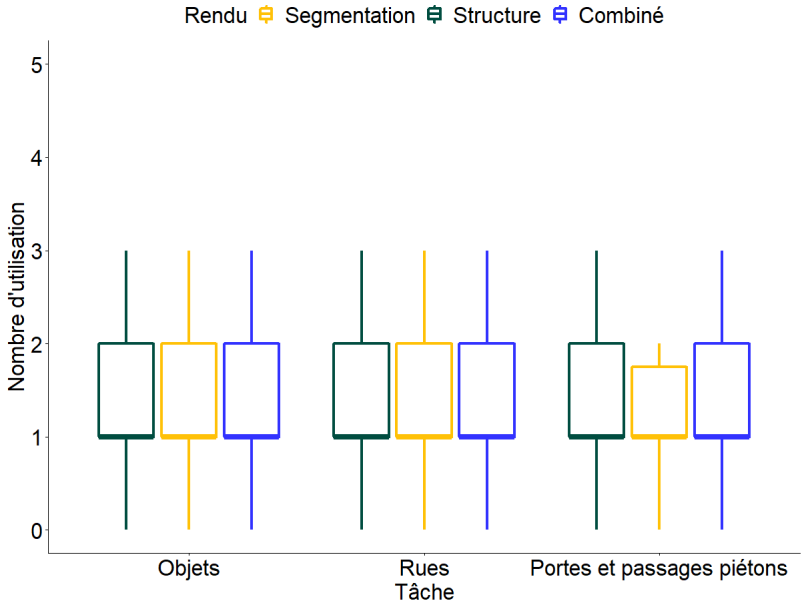


FIGURE 4.13 – Nombre d'utilisation moyens de chacun des trois rendus du mode *Switch* par Tâche. Cette figure montre des résultats non significativement différents. Analyse réalisée sur les 20 sujets.

Une ANOVA à un facteur (Tâche) a montré que le nombre de fois où un rendu est utilisé en dernier est significativement différent selon la tâche ($F(2,171) = 5.914$; $p < 0.001$). Un test post-hoc de Tukey montre que les résultats sont significatifs pour les trois tâches. La FIGURE 4.14 montre des résultats significativement différents et donc un lien entre le dernier rendu utilisé et la tâche réalisée. Le rendu *Combiné* est plus utilisé en dernier que le rendu Structure et le rendu Objets, quelle que soit la tâche réalisée. On note aussi que le rendu Objets est plus utilisé que le rendu Structure pour répondre aux tâches de type *objets*.

4.4 Analyse et discussion des résultats

Nos résultats montrent que la détection des passages piétons et des portes est relativement aisée, indépendamment du rendu utilisé. Ceci n'est pas surprenant, car pour les portes, le rendu consiste systématiquement en un rectangle blanc qui est facile à percevoir. Une étude de Moreno [Moreno et al., 2012] utilisait une méthode différente pour détecter les portes. Leur objectif était différent du nôtre dans le sens où ils cherchaient à donner un indice sonore sur l'emplacement de portes le long de murs dans un couloir. Les passages

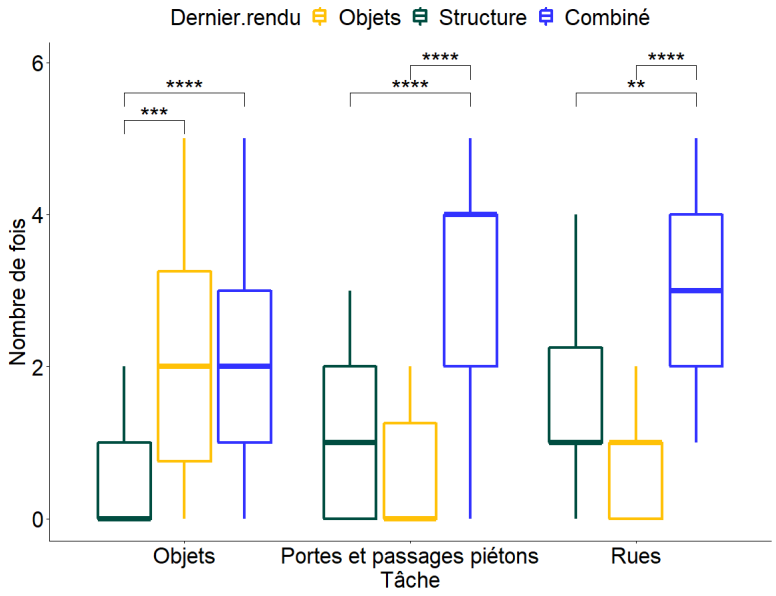


FIGURE 4.14 – Nombre moyen d'utilisation en dernière instance des trois rendus du mode Switch par Tâche. Analyse réalisée sur les 20 sujets.

piétons sont moins faciles à percevoir que les portes, mais plus faciles à percevoir que les autres objets, car ils apparaissent comme une succession de rectangles parallèles.

Par contre, l'identification des objets et la compréhension de l'organisation des rues dépendent fortement du rendu utilisé. Les résultats montrent que le rendu *Switch* est significativement meilleur que le rendu Contrôle et le rendu *Combiné*. Nous avons à nouveau observé que le rendu *Combiné* n'est pas significativement plus efficace que le rendu Contrôle pour la détection des objets, ce qui ne semble pas confirmer les résultats de Sanchez et al. (Sanchez-Garcia et al., 2020). Cette divergence peut s'expliquer par la résolution des implants utilisés dans notre simulateur (21x18) qui est inférieure à celle utilisée dans le leur (32x32). En effet, selon Cha et al. [Cha et al., 1992a], 600 électrodes avec un rendu de type *Scoreboard* (Contrôle) sont suffisantes pour obtenir une perception déjà fonctionnelle. Ces résultats sont également en accord avec ceux du chapitre 3, qui rappelons-le, ont démontré que le rendu *Scoreboard* n'est pas significativement plus efficace que le rendu *Combiné* dans une tâche de compréhension de la scène.

L'analyse des réponses "je ne sais pas" (FIGURE 4.8) montre qu'il y a une différence significative selon les rendus et les tâches. Le rendu Contrôle, obtient plus de réponses "je ne sais pas" que le rendu *Switch* dans les tâches portant sur les objets et les rues. Un des sujets a beaucoup utilisé cette option et a répondu "je ne sais pas" 14 fois pour le mode de rendu *Scoreboard* (contrôle), 6 fois pour le mode de rendu *Combiné* et 7 fois pour le mode *Switch*. Son nombre de bonnes réponses est de 1 pour *Scoreboard*, 6 pour le mode de rendu *Combiné* et 6 pour le mode *Switch*.

En ce qui concerne les temps de réponse, nous obtenons des différences significatives.

Le rendu *Switch* est systématiquement plus utilisé en termes de temps d'utilisation que le rendu Contrôle peu importe la tâche. Il est également plus utilisé que le rendu *Combiné* dans les tâches de type *portes et passages piétons*. Cela peut s'expliquer par le fait que changer de rendu rallonge automatiquement le temps de réponse. En contrepartie, comme le montrent les indices de préférence, la décision est plus sûre et la satisfaction plus grande. On pourrait imaginer qu'avec de l'entraînement, la décision soit prise plus rapidement avec le mode *Switch*. Sur notre courte étude nous n'avons identifié aucun phénomène d'apprentissage au cours de l'expérience, quel que soit le rendu. La FIGURE 4.15 montre l'évolution du temps de réponse moyen en fonction de l'avancée dans les questions et du mode de rendu utilisé.

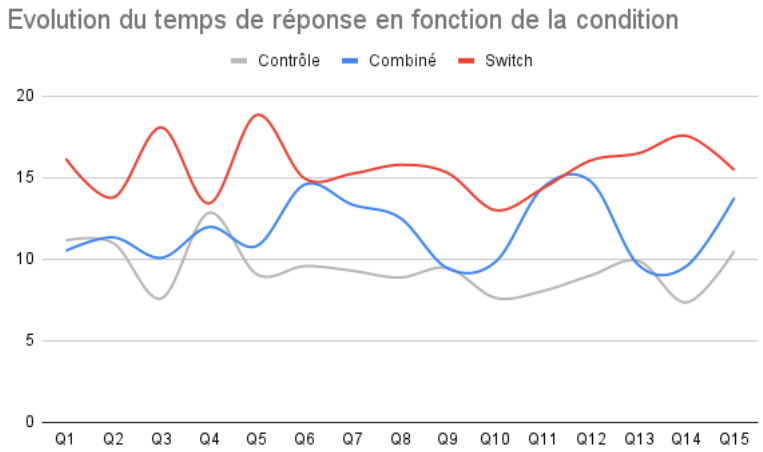


FIGURE 4.15 – Évolution du temps de réponse en fonction de la condition utilisée. L'axe des abscisses représente l'ordre réel des questions posées. L'axe des ordonnées représente le temps de réponse moyen pour ces questions. Les données sont moyennées sur les 20 sujets.

En ce qui concerne l'analyse des jugements subjectifs et du classement des modes de rendu, nous mettons en évidence que le mode de rendu *Switch* présente un net avantage sur le mode de rendu *Scoreboard* dans toutes les catégories. Cela dit, le mode de rendu *Switch* n'est pas significativement préféré par rapport au mode de rendu *Combiné*. En revanche, le classement final montre quand même que le mode de rendu classé premier est le *Switch* devant le mode de rendu *Combiné*.

Les résultats obtenus en regardant de plus près le comportement des sujets lorsqu'ils utilisent le mode de rendu *Switch* sont également intéressants. La plupart des sujets utilisent souvent les rendus dans l'ordre d'affichage des boutons à l'écran. Il semble donc qu'il n'y ait pas de stratégie consciente d'utilisation des rendus de la part des sujets. Par contre, on observe que le dernier rendu choisi est significativement relié avec la tâche réalisée. On observe notamment que les rendus *Combiné* et Objets sont plus souvent les derniers rendus utilisés pour l'identification d'objets. Les rendus *Combiné* et Structure

sont plus souvent les derniers rendus utilisés pour comprendre la configuration des rues. Ceci confirme l'hypothèse selon laquelle les rendus appropriés à chaque tâche sont plus utilisables pour prendre la décision finale. Par contre, le rendu *Combiné* est utilisé plus fréquemment en dernier ce qui va à l'encontre de notre hypothèse sur l'encombrement visuel. En effet, dans ce cas, le rendu Objets devrait être privilégié pour percevoir les objets et le rendu Structure devrait être privilégié pour comprendre les orientations des rues. Cependant, ce résultat est à considérer en fonction du fait que le rendu *Switch* est largement préféré au rendu *Combiné* seul (en dehors de la condition *Switch*, quand il est présenté comme seul rendu disponible). Une interprétation possible est que le rendu *Combiné* est apprécié après avoir utilisé les rendus appropriés pour percevoir les objets dans la scène visuelle. Il permettrait de confirmer la perception unitaire puis d'essayer de les replacer dans la scène globale afin de mieux la comprendre.

4.5 Limitations de l'étude et perspectives

Dans un premier temps, nous pouvons aborder la question de la prédiction des images de segmentation et de structure pour construire les différents rendus. Dans notre étude, ces cartes sont très faciles à obtenir, car les objets de la scène sont labellisés et les caméras paramétrées pour ne capturer que les éléments qui sont pertinents. En condition réelle, il faudrait trouver un autre moyen de récupérer ces informations. L'ajout de réseaux de neurones pourrait permettre de résoudre ce problème. Les deux problèmes majeurs qu'il reste à gérer sont les erreurs dans les prédictions et la vitesse de calcul. Il faut que les prédictions ne soient pas trop loin de la réalité et que ces prédictions soient produites en temps quasi-réel.

Une autre limitation est que notre étude ne permet pas de retranscrire la notion de mouvement étant donné que nous utilisons des images statiques. L'information de mouvement est une information très utile, à tel point que certains dispositifs utilisent des caméras basées événement (caméras événementielles) [Posch et al., 2014] pour créer un rendu phosphénique. Pour résoudre ce problème, tout en prenant en compte la difficulté de charger des modèles 3D sur les ordinateurs des sujets à distance, on pourrait imaginer réaliser la même étude en mettant à la place des clips vidéo.

Pour pousser le réalisme au maximum, il pourrait également être intéressant de réaliser une expérience en conditions réelles. Le dispositif pourrait être composé d'un casque de réalité virtuelle équipé d'une caméra qui filme la scène et d'un boîtier de calcul qui s'occupe de calculer le rendu visuel. Une télécommande permettrait au sujet de changer de rendu en temps réel. Ce dispositif serait une extension du premier proposé par Cha en 1992 [Cha et al., 1992b]. L'intérêt de faire une étude en conditions réelles est double : d'une part, on pourrait proposer un dispositif d'étude complet très proche de la réalité, ce qui permettrait aux futurs travaux d'être menés sur un modèle réaliste et d'autre part, nous

pourrions lors de cette étude mesurer la réaction des sujets dans des tâches de navigation et non plus seulement de perception et de compréhension.

Nous discuterons plus précisément ces perspectives dans le prochain chapitre.

En résumé

Dans ce chapitre, nous avons mis en évidence que fournir un système interactif dans un contexte de perception de scènes soumises à une restitution phosphénique, permet aux sujets d'avoir de biens meilleures performances.

Cette approche répond aux exigences des non-voyants [Ratelle, 2019] d'une part, car elle permet aux sujets de mieux reconnaître les objets/obstacles qui pourraient les gêner dans une tâche de navigation, et d'autre part, car elle permet d'identifier plus facilement les configurations de carrefour. Cette approche est nouvelle et n'a jamais (à notre connaissance) été proposée.

Dans la partie suivante, nous aborderons les perspectives d'amélioration réalisables ainsi que les expériences complémentaires qui nous paraissent pertinentes. Les deux principaux axes d'améliorations sont les suivants : l'amélioration du simulateur (augmentation du réalisme), et l'amélioration des conditions expérimentales.

Conclusion et perspectives

La problématique de nos travaux était de savoir dans quelle mesure il est réalisable de communiquer des informations significatives et utiles à des sujets implantés utilisant un système de neuroprothèse visuelle par l'intermédiaire de systèmes multimodaux et interactifs. Toutes nos démarches sont basées sur l'analyse des besoins des non-voyants établie par Ratelle et Couturier [Ratelle, 2019]. Nous avons pour cela développé un simulateur de vision prothétique aussi proche des conditions réelles que possible, car l'accès à des non-voyants implantés nous a été difficile et que le simulateur nous permet de tester des méthodes non réalisables en conditions réelles (chapitre 2).

Notre première contribution (chapitre 3) a permis de mettre en évidence que l'ajout d'indices sonores dans un contexte de navigation et de compréhension en scène d'intérieur est particulièrement intéressant. Nos résultats significatifs montrent que les sujets arrivent à déterminer les objets présents dans la scène (+18% en moyenne par rapport aux méthodes sans indices audio). Nous montrons également que les objets sont largement mieux placés dans la scène lorsque le dispositif d'aide audio est activé (+31%). Un rendu visuel, aussi élaboré soit-il, a un impact négligeable sur a compréhension et la navigation par des non-voyants comparé à la présence d'indices audio, dans le cas de la simulation d'implants de faible résolution.

Notre deuxième contribution (chapitre 4) montre que donner la possibilité aux sujets de changer en temps réel de méthode de rendu permet d'améliorer significativement leur faculté de compréhension de scènes extérieures. La tâche consistait à identifier un certain nombre d'éléments dans une scène d'extérieur, répartis en trois catégories : *objets*, *intersections de rues* et *portes et passages piétons*. Nous montrons notamment que le mode qui permet de changer en temps réel de méthode de rendu est largement préféré par les utilisateurs.

Pour conclure, nos travaux ont permis de mettre en évidence qu'un rendu seul, c'est-à-dire sans ajout de multimodalité (audio) ou d'interactivité (*Switch*), n'est pas suffisant pour réaliser des tâches simples de compréhension et de navigation, et cela, même avec un implant simulé loin des conditions réelles de perception. Les prochains objectifs de la recherche dans le domaine de la restitution visuelle prothétique devraient donc probablement se concentrer sur l'interaction et sur la multimodalité.

Nous présentons maintenant les perspectives liées aux travaux présentés jusqu'à mainte-

nant. Nous aborderons les améliorations possibles en ce qui concerne la simulation du rendu phosphénique (1) et des conditions expérimentales (2). Puis, nous proposons quelques études complémentaires qui permettraient de renforcer nos résultats (3).

1 Amélioration du rendu phosphénique

Nos expérimentations ont été conduites sur des sujets voyants par l'intermédiaire du simulateur décrit dans le chapitre 2, l'accès à un échantillon de non-voyants implantés étant difficile. La question n'est pas tant celle de l'applicabilité directe des résultats obtenus sur des sujets voyants aux non-voyants que celle de la validité de la démarche. Notre objectif est davantage d'évaluer à quel point des méthodes de rendus sophistiquées, en supposant que les implants futurs permettent de résoudre les problèmes liés à la vision prothétique, permettent aux patients de résoudre des tâches du quotidien.

Notre simulateur permet de réaliser des expérimentations pour des tâches de compréhension/reconnaissance ou de navigation. Cependant, il y a des défauts que notre simulateur ne modélise pas. Dans la suite de cette section, nous proposons des pistes d'amélioration du simulateur pour se rapprocher des conditions réelles de perceptions. Nous proposons plusieurs améliorations possibles concernant la représentation du rendu phosphénique présentée au chapitre 2, et, plus précisément, nous discutons quatre problèmes partiellement ou non traités : évanescence, persistance, géométrie des phosphènes et apparition multiple.

1.1 Phénomène d'évanescence

Nous rappelons que le phénomène d'évanescence apparaît dès qu'une électrode reçoit une séquence de stimuli. Quand cela se produit, la perceptibilité du phosphène associé décroît. Pour intégrer ce phénomène au simulateur, il est possible de pondérer l'intensité de stimulation du phosphène considéré par une exponentielle décroissante. Ainsi, pour chaque électrode simulée, on aurait deux paramètres, le premier serait l'intensité maximale de restitution de cette électrode, et le second serait le temps qu'il faut à l'électrode pour "récupérer son plein potentiel". La FIGURE 4.1 illustre cette proposition.

1.2 Phénomène de persistance

Pour rappel, le phénomène de persistance se produit dès lors qu'une électrode passe de l'état stimulé à non stimulé. En réalité, la perception du phosphène lié à cette électrode ne s'arrête pas d'un coup, comme nous l'avons déjà dit dans le chapitre 1. Pour simuler ce phénomène, nous pourrions également pondérer l'intensité du phosphène par une fonction strictement décroissante lorsqu'il n'y a plus de stimulation de l'électrode associée. On

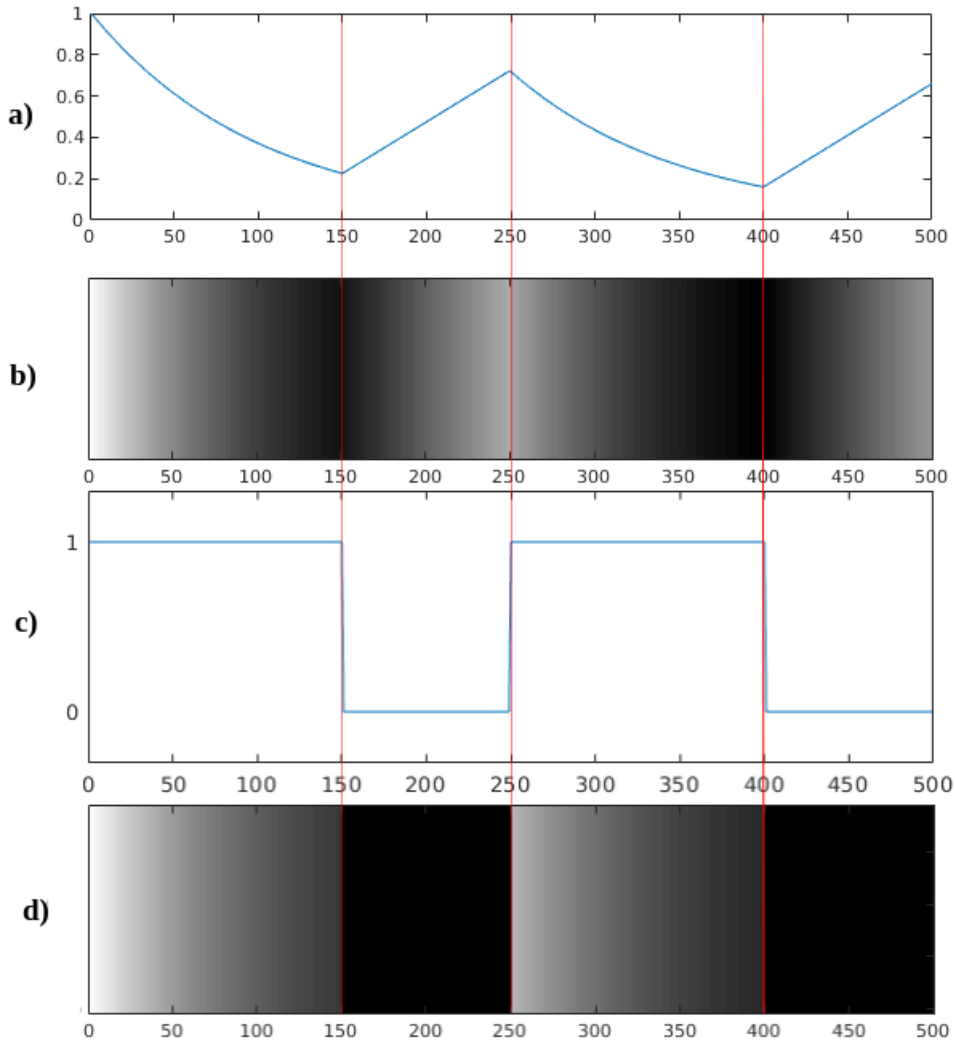


FIGURE 4.1 – Cette figure illustre sous la forme de 4 graphiques l'évolution de l'intensité d'un phosphène perçue, pondéré par une exponentielle décroissante. L'intensité maximale de stimulation correspond à la perception d'un phosphène blanc pur (255 en niveau de gris). On considère que l'électrode est stimulée à haute fréquence. La figure a) montre l'évolution de l'intensité du phosphène en fonction du temps. Au départ, l'intensité est maximale, puis décroît de façon exponentielle. La figure b) est une représentation théorique de l'intensité du phosphène sous forme de gradient de niveau de gris. La figure c) représente les intervalles durant lesquels l'électrode est stimulée. Dans notre cas, elle est stimulée de 0 à 150 et de 250 à 400. En 150, on arrête de stimuler l'électrode, ce qui se traduit dans la figure a) par une recharge linéaire de l'intensité. La figure d) montre l'intensité réelle perçue. En 150, comme on arrête de stimuler l'électrode, il n'y a plus de phosphènes perçus. Notons toutefois qu'ici, nous ne prenons pas en compte le phénomène de persistance. Sur toutes les figures, l'axe des abscisses représente des unités de temps.

aurait dans ce cas un paramètre variable pour chaque électrode qui est cette fonction de décroissance. La FIGURE 4.2 illustre notre proposition.

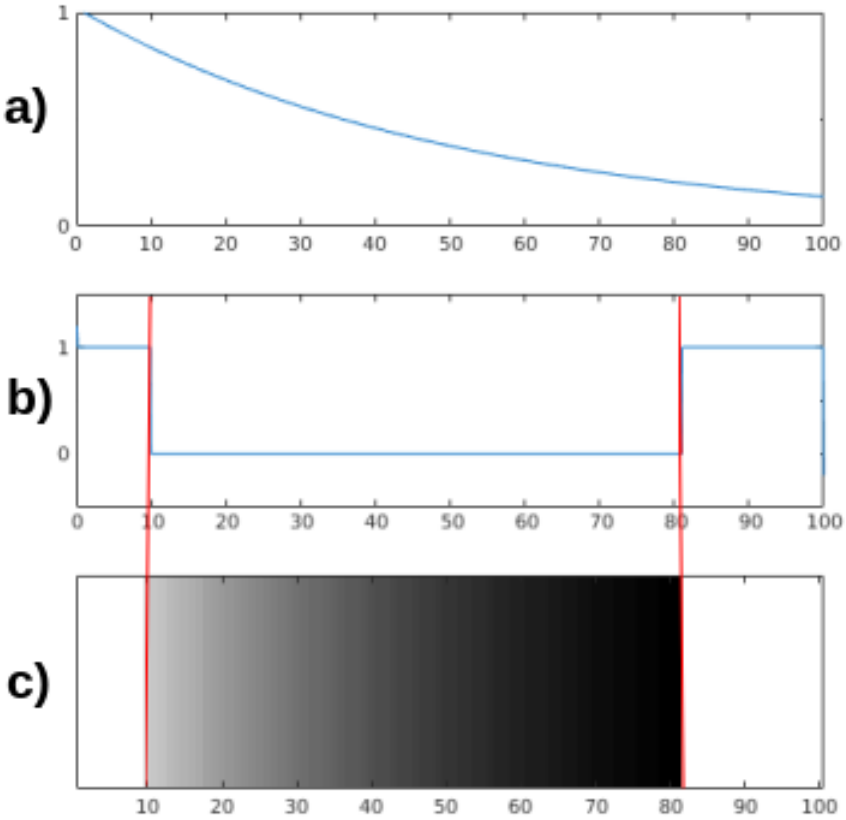


FIGURE 4.2 – Cette figure illustre la mise en place du phénomène de persistance phosphénique. L'intensité du phosphène ne passe pas de 1 à 0 dès que la stimulation de l'électrode s'arrête. La figure a) montre la fonction de décroissance de l'intensité du phosphène. La figure b) montre les moments où l'électrode est stimulée (entre 0 et 10 – 80 et 100). La figure c) montre comment l'intensité du phosphène varie au cours du temps. Sur cette figure, le phénomène d'évanescence n'est pas simulé. En situation réelle, il faudrait simuler les deux en même temps. Notons que la décroissance de la fonction utilisée est plus rapide que la fonction utilisée dans l'évanescence.

1.3 Géométrie des phosphènes

Notre simulateur ne permet pour l'instant de modéliser que des phosphènes de forme ronde. Pour se rapprocher des perceptions décrites par les patients implantés en essais cliniques, il faut intégrer des formes de phosphènes plus complexes et variables. Nous proposons ici deux axes d'améliorations possibles, plus réalistes, le premier est de décrire les phosphènes sous la forme d'hexagones, le second est de les décrire sous forme d'ellipses (ce qui permettrait de simuler, par exemple, de long traits fins, comme cela a été rapporté

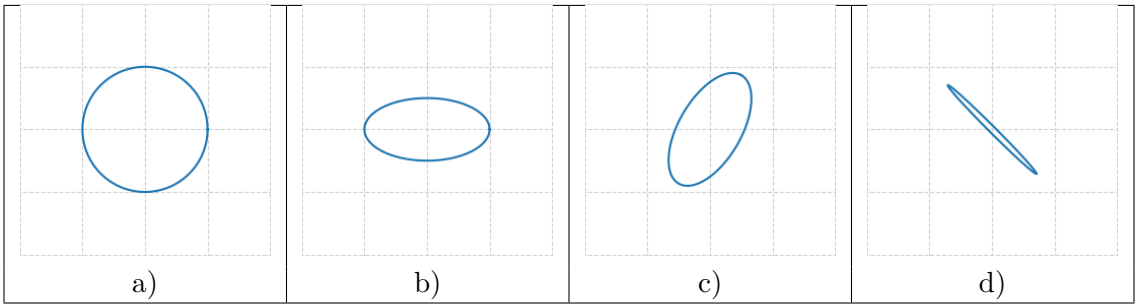


TABLEAU 4.7 – Différentes géométries pour les phosphènes. L'image a) montre une ellipse circulaire (cercle), l'image b) montre une ellipse aplatie sur l'axe des ordonnées, l'image c) montre une ellipse aplatie qui a subi une rotation et l'image d) montre une ellipse très aplatie qui a subi une rotation. L'image d) se rapproche fortement de certaines descriptions de phosphènes provenant de patients implantés.

par certains patients implantés) [Nanduri et al., 2008].

Dans le simulateur, pour déterminer si un pixel doit être allumé dans le rendu final, nous calculons sa distance au centre du cercle représentant le phosphène. Si la distance entre ces deux points est inférieure au rayon du phosphène, on attribue alors une intensité non-nulle au pixel. En revanche, si la distance est plus grande que le rayon, l'intensité du pixel est de zéro. Ce procédé peut être généralisé aux hexagones (et même plus généralement à n'importe quel polygone qui possède au moins trois sommets). On a ainsi une méthode caractérisant si un point (et donc un pixel) est à l'intérieur de la zone à rendre visible.

Pour remplacer les cercles par des ellipses, il faut passer de trois paramètres (coordonnées du centre du cercle (x, y) et rayon) à six paramètres (cordonnées des deux foyers, demi grand axe et rotation). Il est également possible de caractériser le fait qu'un point est dans une ellipse. De manière analogue au cas des hexagones et des cercles, il suffit de donner une intensité non-nulle aux pixels à l'intérieur de l'ellipse. Le tableau 4.7 montre différentes géométries de phosphènes qui pourraient ainsi être construits.

1.4 Apparition de phosphènes multiples

Parfois, lors de la stimulation d'une électrode, cette dernière ne produit pas l'apparition du phosphène attendu, mais peut faire apparaître un ou plusieurs phosphènes normalement générés par la stimulation d'électrodes voisines. Pour simuler ce phénomène, il faut introduire une probabilité de mauvaise apparition (relativement faible, de l'ordre de 1%) et propre à chaque électrode. Dans ce cas il y a donc une chance sur cent que l'électrode ne produise pas l'apparition du phosphène désiré. Ensuite, il faut déterminer, parmi les électrodes voisines, lesquelles seront faussement stimulées. On peut par exemple choisir, le nombre d'électrodes stimulées (entre une et quatre en 4-connectivité) de manière uniforme.

2 Amélioration des conditions expérimentales

À cause de la crise sanitaire qui a démarré en France au début de l'année 2020, les expériences que nous avions commencé à mettre en place ont rapidement été mises de côté. Les deux expériences conduites se sont déroulées à distance. Cette section a pour objectif de présenter quelques pistes d'amélioration des conditions expérimentales, que ce soit en termes de réalisme ou de confort d'utilisation. Nous proposons de discuter de l'ajout d'un système de suivi du regard (*eye-tracking* en anglais), de l'amélioration du dispositif d'interaction et de l'ajout de la notion de mouvement réel en intégrant un dispositif de réalité virtuelle.

2.1 Module de suivi du regard

En conditions réelles, les non-voyants ne peuvent pas découvrir leur environnement en observant avec leurs yeux. En effet, le module d'acquisition est très souvent une caméra placée sur des lunettes. Ainsi, ils sont obligés de déplacer toute leur tête pour observer (explorer et comprendre) l'environnement. De plus, l'implant est fixe sur le site d'implantation. Cela signifie que s'ils gardent la tête fixe, ils ne peuvent pas non plus observer les différents endroits de la restitution visuelle qui leur est proposée. Leur vision est centrée sur le milieu de la restitution, et donc les informations de périphéries sont plus difficiles à lire. Dans notre simulateur, les sujets peuvent librement observer toute la restitution visuelle.

L'ajout d'un système de suivi du regard permettrait de résoudre ce problème. Une fois les déplacements des yeux calculés, il suffit de déplacer la zone de rendu proportionnellement à ce déplacement. La Figure 4.3 illustre la manière dont le déplacement serait fait dans notre simulateur qui effectue le rendu sur un écran.

Il y a néanmoins trois problèmes avec cet ajout. Premièrement, les yeux sont en constant mouvement. Ce qui signifie que la zone de rendu va subir un effet de tremblement permanent pour les sujets en simulation. Les non-voyants implantés ne subissent pas du tout ce phénomène puisque leur tête n'est pas systématiquement en train de bouger. Il faut donc probablement introduire un seuil de déplacement à partir duquel on considère que les yeux n'ont pas trop bougé. Deuxièmement, le cerveau n'est pas habitué à ce genre de processus. Il est possible de rendre le système d'orientation des sujets en simulation confus et de créer des nausées et maux de têtes. Enfin, il faut s'assurer que le taux de rafraîchissement du déplacement soit au moins de 30 images par seconde pour garantir le caractère temps réel du processus.

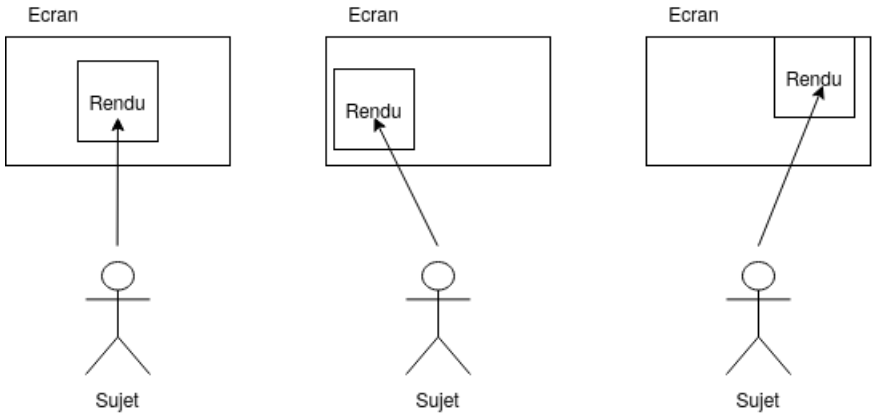


FIGURE 4.3 – Illustration du déplacement de la zone de rendu en fonction du regard du sujet. L'extrémité de la flèche noire correspond à la zone de l'écran regardée par le sujet. Le déplacement de la zone de rendu se fait en temps réel au fur et à mesure que le regard se déplace.

2.2 Utilisation d'un casque de réalité virtuelle - entre simulation et environnement réel

Notre système d'interaction est relativement simple étant donné que les sujets se trouvent systématiquement face à leur écran pour prendre part aux expériences. Les interactions se font principalement au clavier et à la souris. L'interface proposée aux sujets a été longuement étudiée. La disposition des différents éléments se veut intuitive et simple d'accès, ce qu'a confirmé la majorité des sujets de tests.

Cependant, cette approche de la simulation reste loin des conditions réelles d'utilisation. Une amélioration que nous proposons pour augmenter le degré de réalisme des expériences est de passer d'un mode assis et passif (utilisation du clavier et de la souris) à un mode debout et actif (exploration de l'environnement avec des déplacements réels). Au tout début de ces recherches, l'objectif était de développer un tel système expérimental. Malheureusement, la crise sanitaire a balayé cette possibilité du fait de l'impossibilité de conduire l'expérience en laboratoire. Un prototype avait été initialement développé pour réaliser ce genre d'expériences. Il était constitué d'un casque HTC Vive Pro Eye (FIGURE 4.4) relié par un câble à un ordinateur qui s'occupait de réaliser la simulation de l'environnement. Les tâches modélisées étaient des tâches de suivi de ligne blanche au sol comme cela a pu être fait auparavant dans la littérature.

L'avantage de ce casque est qu'il permet de faire de l'acquisition de scène réelle grâce à deux caméras placées à l'avant. Grâce à elles, il est possible de ne plus simuler d'environnement, mais de réaliser des tâches en situation réelle, c'est-à-dire avec toutes les possibilités qu'offre la réalité. Dernièrement, nous avons également fait l'acquisition du HTC Vive Flow (FIGURE 4.5) qui est sans fil, et qui devrait permettre de réaliser ces ex-



FIGURE 4.4 – Illustration du HTC Vive Pro Eye.

périences aussi bien en intérieur qu'en extérieur. Le casque HTC Vive Flow dispose aussi de caméras à l'avant du casque.



FIGURE 4.5 – Illustration du HTC Vive Flow.

Il faut toutefois être en mesure de calculer les rendus dans ces systèmes légers. C'est ici qu'interviennent les modèles de réseaux de neurones embarqués, qui assurent les prédictions (extraction de caractéristiques) en temps réel sur des dispositifs moins puissants.

2.3 Intégration des réseaux de neurones et validation des résultats en cas réel

À l'heure actuelle, dans le simulateur utilisé, les cartes de segmentation sémantique et d'extraction de structure sont construites sur la base d'une labellisation des objets dans la scène. Une amélioration possible réside dans l'intégration d'un module de prédiction des cartes de segmentation sémantique et de structure via l'utilisation de réseaux de neurones. Deux questions se posent alors : comment entraîner et intégrer les réseaux au simulateur et comment évaluer la pertinence de l'utilisation et des résultats (temps de calculs et erreurs de prédiction).

Entraînement et intégration

En ce qui concerne l'entraînement, il est possible de réaliser ce qu'on appelle de la spécification (*fine-tuning* en anglais). C'est un procédé qui consiste à prendre un réseau déjà entraîné à bien résoudre une tache et de le spécialiser sur des données plus précises.

Pour l'intégration, il y a deux cas : la simulation et les conditions réelles. En simulation, il faut établir un canal de communication entre le module de simulation et le module de calcul. Cela se fait facilement en utilisant un système de sockets. En conditions réelles, il faut que le réseau puisse être embarqué sur le dispositif de calcul (1.5.2).

Évaluation de la pertinence et des résultats

Lors de l'utilisation de réseaux de neurones embarqués, il y a un critère très important à respecter. Il faut que la prédiction puisse être produite à une fréquence suffisamment élevée pour être en temps réel. Dans notre cas, c'est très important que les représentations phosphéniques soient générées assez régulièrement pour que les potentiels obstacles se trouvant sur l'itinéraire du patient implanté soient rendus à temps.

Le dernier élément à prendre en compte est les erreurs de prédiction que peuvent produire les réseaux de neurones. Dans le cas d'une segmentation sémantique, une erreur de classification d'un pixel aura peu d'impact sur la restitution phosphénique engendrée. En revanche, une erreur dans la reconnaissance d'un objet peut avoir des conséquences dramatiques (par exemple si l'information produite est "table" au lieu de "camion"). Il est donc nécessaire d'évaluer à quel point les réseaux sont fiables en termes de prédiction et de ne considérer que ceux qui respectent un taux de fiabilité suffisant.

De plus les prédictions sont généralement réalisées image par image, les questions de l'instabilité des prédictions et de l'impact que cela peut avoir sur le confort visuel se posent alors.

3 Autres expériences réalisables

Dans cette section, nous présentons quelques études complémentaires qui peuvent renforcer nos résultats ou apporter de nouvelles contributions.

3.1 Switch + Audio

Il serait intéressant dans un premier temps de coupler les résultats montrés dans les chapitres 3 et 4 et d'analyser les performances des sujets dans des tâches de navigation. Cette étude permettrait de mettre en compétition et en coopération la possibilité de changer de rendu en temps réel et la présence d'indices audio.

L'intérêt est ici de voir si une des deux techniques est plus utilisée par les sujets ou si la combinaison des deux permet d'améliorer encore plus les performances.

3.2 Encombrement

Nos environnements simulés sont très peu chargés en objets. D'une part, dans l'expérience du chapitre 3, nous ne présentons que huit objets assez larges dans une scène composée de quatre grandes pièces. D'autre part, dans le chapitre 4, très peu d'objets sont modélisés et utilisés (environ une dizaine).

Il serait intéressant de voir comment évoluent les performances des sujets lorsque l'encombrement des scènes augmente. Par exemple, dans un bureau classique, on retrouve de nombreux objets tels que des stylos, des claviers, des souris, des livres, des lampes, des sacs, etc. De même, en situation extérieure, on peut trouver des barrières, des plots de circulations, des vélos, etc.

Analyser les performances des sujets dans le même genre d'expérimentations que celles réalisées dans les chapitres 3 et 4 permettrait de valider nos résultats lorsque de nombreux objets sont présents dans la scène et donc que la scène s'approche plus de la réalité. Notre hypothèse est dans ce cas que le rendu combiné devrait être bien meilleur que le rendu scoreboard dans ce type d'environnement.

3.3 Apprentissage par renforcement

À l'origine, une des idées de cette thèse était de construire un système apprenant à construire la meilleure restitution visuelle possible en temps réel. Un tel système est basé sur ce qu'on appelle l'apprentissage par renforcement, dont la FIGURE 4.6 illustre le fonctionnement. Les différents acteurs de l'environnement sont le système et l'agent. L'agent perçoit un état communiqué par le système. Sur la base de cet état, il effectue une action sur le système qui va le conduire à changer d'état et à générer une récompense. L'objectif de l'agent est de réussir à trouver la séquence d'actions qui va maximiser les récompenses au

fur et à mesure de l'apprentissage. Cette technique se décline avec les réseaux de neurones pour résoudre des problèmes de grande dimension.

Dans notre cas, on pourrait modéliser le système par l'ensemble des éléments décrivant la scène, l'image capturée par la caméra, la position et l'orientation du sujet dans la simulation, les dernières décisions prises par le sujet, etc. La difficulté est de savoir quand attribuer quelle récompense. Bien entendu, ce choix dépend fortement de la tâche considérée.

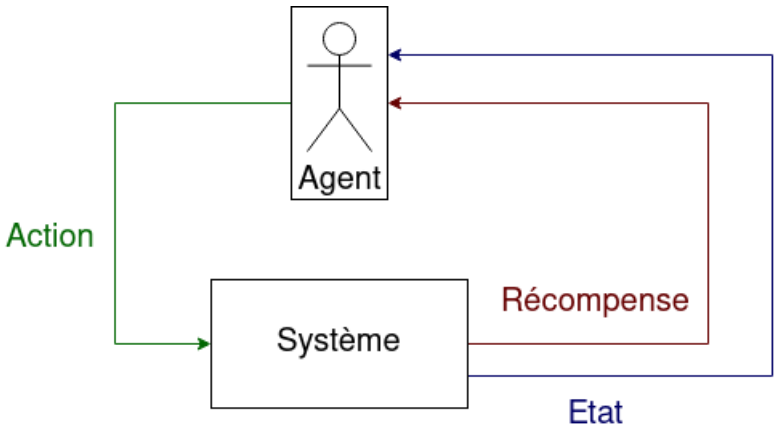


FIGURE 4.6 – Schéma de fonctionnement de l'apprentissage par renforcement. L'agent effectue des actions sur le système ce qui le fait changer d'état et génère une récompense.

Une amélioration possible d'un tel procédé est d'intégrer pendant l'apprentissage le retour d'un agent superviseur humain. L'intérêt d'un tel ajout est double, on peut limiter l'exploration d'actions qui ont peu d'intérêt pendant l'apprentissage (par exemple n'utiliser que 2 électrodes pour générer une restitution), et on peut également adapter un algorithme entraîné au sujet pendant l'utilisation grâce à ses retours (par exemple, si le sujet trouve qu'il ne peut pas prendre de décision, il presse un bouton qui informe le réseau que la restitution n'est pas suffisante et le réseau se met à jour en fonction de cette information). C'est ce qu'on appelle "apprentissage par renforcement avec humain dans la boucle". Les questions qui se posent alors sont l'intégration d'un module de réseaux de neurones dans le simulateur et le temps nécessaire pour converger vers des prédictions stables.

Annexe 1

Questionnaire MeCue utilisé dans l'étude portant sur l'ajout d'indices audio (étude 1)

— Le rendu est parfaitement adapté à la tâche que je viens de réaliser.
— Je considère ce rendu comme extrêmement efficace.
— A l'aide de ce rendu, j'ai pu atteindre les objectifs de la tâche.
— Le rendu est facile à utiliser.
— On comprend rapidement comment utiliser le rendu.
— L'utilisation du rendu est facile à comprendre.
— Le rendu est plaisant.
— Le rendu est attrayant.
— Le rendu est agréable.
— Si j'étais un non-voyant implanté, je voudrais absolument ce rendu.
— Si j'étais un non-voyant implanté, ce rendu serait très important pour moi.
— Si j'étais un non-voyant implanté et que je perdais ce rendu, je serais dévasté.
— Ce rendu me plaît énormément.
— L'utilisation de ce rendu me rassure.
— Quand j'utilise ce rendu, je me sens très satisfait.
— Quand j'utilise ce rendu, je me sens très détendu.
— Ce rendu me rend heureux.
— L'utilisation de ce rendu m'apaise.
— Ce rendu m'énerve.
— Ce rendu me fatigue.
— Ce rendu me met en colère.
— Ce rendu me frustre.
— Ce rendu me déçoit.
— Je n'ai aucune motivation à utiliser ce rendu.
— Je n'échangerai ce rendu contre aucun autre.
— Ce rendu à l'air très perfectionné.
— Si j'étais un non-voyant implanté, je n'hésiterais pas à choisir ce rendu.

— Si j'étais un non-voyant implanté, j'utiliserais ce rendu chaque jour.
— Si j'étais un non-voyant implanté, je serais impatient de pouvoir utiliser ce rendu.
— Quand j'utilise ce rendu, je ne vois pas le temps passer.
— Avec ce rendu, je perçois bien les objets dans la scène.
— Avec ce rendu, je comprends parfaitement la configuration de la scène.
— Avec ce rendu, je peux me déplacer sans problèmes.
— Avec ce rendun j'arrive à éviter les obstacles sur mon chemin.
— Comment évalueriez-vous ce rendu dans son ensemble ?

Annexe 2

Questions Likert pour l'étude sur l'intérêt d'un rendu interactif (étude 2)

— J'ai utilisé ce mode de restitution avec plaisir.
— J'ai eu des difficultés à utiliser ce mode de restitution.
— Ce mode de restitution est adapté à la tâche que je devais résoudre.
— Ce mode de restitution me procure suffisamment d'informations pour réussir la tâche soumisse.
— (Uniquement dans le cas du mode *Switch*) : J'ai trouvé la possibilité de changer de rendu très utile.

Bibliographie

[Abbott et al., 2018] Abbott, C. J., Nayagam, D. A., Luu, C. D., Epp, S. B., Williams, R. A., Salinas-LaRosa, C. M., Villalobos, J., McGowan, C., Shivdasani, M. N., Burns, O., et al. (2018). Safety studies for a 44-channel suprachoroidal retinal prosthesis : a chronic passive study. *Investigative Ophthalmology & Visual Science*, 59(3):1410–1424.

[Ayton et al., 2014] Ayton, L. N., Blamey, P. J., Guymer, R. H., Luu, C. D., Nayagam, D. A., Sinclair, N. C., Shivdasani, M. N., Yeoh, J., McCombe, M. F., Briggs, R. J., et al. (2014). First-in-human trial of a novel suprachoroidal retinal prosthesis. *PloS one*, 9(12):e115239.

[Baheti et al., 2019] Baheti, B., Gajre, S., and Talbar, S. (2019). Semantic scene understanding in unstructured environment with deep convolutional neural network. In *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)*, pages 790–795. IEEE.

[Bloch et al., 2019] Bloch, E., Luo, Y. H.-L., and da Cruz, L. (2019). Advances in retinal prosthesis systems. *Therapeutic Advances in Ophthalmology*, 11.

[Bourne et al., 2017] Bourne, R. R., Flaxman, S. R., Braithwaite, T., Cicinelli, M. V., Das, A., Jonas, J. B., Keeffe, J., Kempen, J. H., Leasher, J., Limburg, H., et al. (2017). Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment : a systematic review and meta-analysis. *The Lancet Global Health*, 5(9):e888–e897.

[Boyle et al., 2002] Boyle, J., Maeder, A., and Boles, W. (2002). Image enhancement for electronic visual prostheses. *Australasian Physics & Engineering Sciences in Medicine*, 25(2):81–86.

[Bradley et al., 2005] Bradley, D. C., Troyk, P. R., Berg, J. A., Bak, M., Cogan, S., Erickson, R., Kufta, C., Mascaro, M., McCreery, D., Schmidt, E. M., et al. (2005). Visuotopic mapping through a multichannel stimulating implant in primate v1. *Journal of neurophysiology*, 93(3):1659–1670.

[Brindley and Lewin, 1968] Brindley, G. S. and Lewin, W. S. (1968). The sensations produced by electrical stimulation of the visual cortex. *The Journal of physiology*, 196(2):479–493.

[Calabrèse, 2011] Calabrèse, A. (2011). *Désorganisation fonctionnelle des systèmes oculo-*

moteur et visuo-attentionnnel chez les patients basse vision :-Approche psychophysique. PhD thesis, Aix-Marseille 2.

[Cha et al., 1992a] Cha, K., Horch, K., and Normann, R. A. (1992a). Simulation of a phosphene-based visual field : visual acuity in a pixelized vision system. *Annals of biomedical engineering*, 20(4):439–449.

[Cha et al., 1992b] Cha, K., Horch, K. W., and Normann, R. A. (1992b). Mobility performance with a pixelized vision system. *Vision research*, 32(7):1367–1372.

[Chai et al., 2008] Chai, X., Li, L., Wu, K., Zhou, C., Cao, P., and Ren, Q. (2008). C-sight visual prostheses for the blind. *IEEE Engineering in Medicine and Biology Magazine*, 27(5):20–28.

[Chen et al., 2017] Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. (2017). Deeplab : Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848.

[Chow et al., 2010] Chow, A. Y., Bittner, A. K., and Pardue, M. T. (2010). The artificial silicon retina in retinitis pigmentosa patients (an american ophthalmological association thesis). *Transactions of the American Ophthalmological Society*, 108:120.

[Chow et al., 2004] Chow, A. Y., Chow, V. Y., Packo, K. H., Pollack, J. S., Peyman, G. A., and Schuchard, R. (2004). The artificial silicon retina microchip for the treatment of visionloss from retinitis pigmentosa. *Archives of ophthalmology*, 122(4):460–469.

[Conradie et al., 2014] Conradie, P., Mioch, T., and Saldien, J. (2014). Blind user requirements to support tactile mobility. *Tactile Haptic User Interfaces for Tabletops and Tablets (TacTT 2014)*, pages 48–53.

[Cun, 1986] Cun, Y. L. (1986). Learning process in an asymmetric threshold network. In *Disordered systems and biological organization*, pages 233–240. Springer.

[da Cruz et al., 2016] da Cruz, L., Dorn, J. D., Humayun, M. S., Dagnelie, G., Handa, J., Barale, P.-O., Sahel, J.-A., Stanga, P. E., Hafezi, F., Safran, A. B., et al. (2016). Five-year safety and performance results from the argus ii retinal prosthesis system clinical trial. *Ophthalmology*, 123(10):2248–2254.

[Dagnelie et al., 2017] Dagnelie, G., Christopher, P., Arditi, A., da Cruz, L., Duncan, J. L., Ho, A. C., Olmos de Koo, L. C., Sahel, J.-A., Stanga, P. E., Thumann, G., et al. (2017). Performance of real-world functional vision tasks by blind subjects improves after implantation with the argus® ii retinal prosthesis system. *Clinical & experimental ophthalmology*, 45(2):152–159.

[Dagnelie et al., 2007] Dagnelie, G., Keane, P., Narla, V., Yang, L., Weiland, J., and Humayun, M. (2007). Real and virtual mobility performance in simulated prosthetic vision. *Journal of neural engineering*, 4(1):S92.

[Dagnelie et al., 2001] Dagnelie, G., Thompson, R. W., Baraett, G. D., and Zhang, W. (2001). Simulated prosthetic vision : perceptual and performance measures. In *Vision Science and its Applications*, page FC2. Optica Publishing Group.

[Dagnelie et al., 2006] Dagnelie, G., Walter, M., and Yang, L. (2006). Playing checkers : detection and eye–hand coordination in simulated prosthetic vision. *Journal of Modern Optics*, 53(9):1325–1342.

[Delbeke et al., 2003] Delbeke, J., Oozeer, M., and Veraart, C. (2003). Position, size and luminosity of phosphenes generated by direct optic nerve stimulation. *Vision research*, 43(9):1091–1102.

[Denis, 2014] Denis, G. (2014). *Apport de la vision par ordinateur dans l'utilisabilité des neuroprothèses visuelles*. PhD thesis, Toulouse 3.

[Denis et al., 2014] Denis, G., Macé, M. J.-M., and Jouffrais, C. (2014). Simulated prosthetic vision : object recognition and localization approach. In *6ème réunion annuelle de l'ITMO" Technologies pour la santé" 2014*.

[Dias et al., 2015] Dias, D. et al. (2015). A sensor platform for the visually impaired to walk straight avoiding obstacles. In *2015 9th International Conference on Sensing Technology (ICST)*, pages 838–843. IEEE.

[Dowling et al., 2005] Dowling, J., Boles, W., and Maeder, A. (2005). Mobility assessment using simulated arti. cial human vision. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-Workshops*, pages 32–32. IEEE.

[Duret et al., 2006] Duret, F., Brelén, M. E., Lambert, V., Gérard, B., Delbeke, J., and Veraart, C. (2006). Object localization, discrimination, and grasping with the optic nerve visual prosthesis. *Restorative neurology and neuroscience*, 24(1):31–40.

[Echenique and Graffigna, 2011] Echenique, A. and Graffigna, J. (2011). Electrical stimulation of mechanoreceptors. In *Journal of Physics : Conference Series*, volume 332, page 012044. IOP Publishing.

[Eckmiller et al., 2005] Eckmiller, R., Neumann, D., and Baruth, O. (2005). Tunable retina encoders for retina implants : why and how. *Journal of neural engineering*, 2(1):S91.

[Fan et al., 2014] Fan, M. Y., Bao, J. T., and Tang, H. R. (2014). A guide cane system for assisting the blind in travelling in outdoor environments. In *Applied Mechanics and Materials*, volume 631, pages 568–571. Trans Tech Publ.

[Farvardin et al., 2018] Farvardin, M., Afarid, M., Attarzadeh, A., Johari, M. K., Mehryar, M., Nowroozzadeh, M. H., Rahat, F., Peyvandi, H., Farvardin, R., and Nami, M. (2018). The argus-ii retinal prosthesis implantation ; from the global to local successful experience. *Frontiers in neuroscience*, 12:584.

[Fernandez, 2018] Fernandez, E. (2018). Development of visual neuroprostheses : trends and challenges. *Bioelectronic medicine*, 4(1):1–8.

[Fernández et al., 2014] Fernández, E., Greger, B., House, P. A., Aranda, I., Botella, C., Albisua, J., Soto-Sánchez, C., Alfaro, A., and Normann, R. A. (2014). Acute human brain responses to intracortical microelectrode arrays : challenges and future prospects. *Frontiers in neuroengineering*, 7:24.

[Fernández and Normann, 2017] Fernández, E. and Normann, R. A. (2017). Cortivis approach for an intracortical visual prostheses. In *Artificial vision*, pages 191–201. Springer.

[Fernandez-Labrador et al., 2018] Fernandez-Labrador, C., Facil, J. M., Perez-Yus, A., Demonceaux, C., and Guerrero, J. J. (2018). Panoroom : From the sphere to the 3d layout. *arXiv preprint arXiv:1808.09879*.

[Fornos et al., 2012] Fornos, A. P., Sommerhalder, J., da Cruz, L., Sahel, J. A., Mohand-Said, S., Hafezi, F., and Pelizzone, M. (2012). Temporal properties of visual perception on electrical stimulation of the retina. *Investigative ophthalmology & visual science*, 53(6):2720–2731.

[Fornos et al., 2008] Fornos, A. P., Sommerhalder, J., Pittard, A., Safran, A. B., and Pelizzone, M. (2008). Simulation of artificial vision : Iv. visual information required to achieve simple pointing and manipulation tasks. *Vision research*, 48(16):1705–1718.

[Gabel, 2017] Gabel, V. P. (2017). *Artificial vision*. Springer.

[Ghiasi et al., 2019] Ghiasi, G., Lin, T.-Y., and Le, Q. V. (2019). Nas-fpn : Learning scalable feature pyramid architecture for object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7036–7045.

[Grumet et al., 2000] Grumet, A. E., Wyatt Jr, J. L., and Rizzo III, J. F. (2000). Multi-electrode stimulation and recording in the isolated retina. *Journal of neuroscience methods*, 101(1):31–42.

[Gupta et al., 2015] Gupta, S., Sharma, I., Tiwari, A., and Chitranshi, G. (2015). Advanced guide cane for the visually impaired people. In *2015 1st International Conference on Next Generation Computing Technologies (NGCT)*, pages 452–455. IEEE.

[Hao et al., 2020] Hao, S., Zhou, Y., and Guo, Y. (2020). A brief survey on semantic segmentation with deep learning. *Neurocomputing*, 406:302–321.

[Hayes et al., 2003] Hayes, J. S., Yin, V. T., Piyathaisere, D., Weiland, J. D., Humayun, M. S., and Dagnelie, G. (2003). Visually guided performance of simple tasks using simulated prosthetic vision. *Artificial organs*, 27(11):1016–1028.

[He et al., 2016] He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

[He et al., 2017] He, L., Lee, K., Lewis, M., and Zettlemoyer, L. (2017). Deep semantic role labeling : What works and what's next. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1 : Long Papers)*, pages 473–483.

[Hebb, 1949] Hebb, D. O. (1949). The first stage of perception : growth of the assembly. *The Organization of Behavior*, 4:60–78.

[Hornig et al., 2007] Hornig, R., Zehnder, T., Velikay-Parel, M., Laube, T., Feucht, M., and Richard, G. (2007). The imi retinal implant system. In *Artificial sight*, pages 111–128. Springer.

[Howard et al., 2017] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H. (2017). Mobilenets : Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.

[Hu et al., 2014] Hu, J., Xia, P., Gu, C., Qi, J., Li, S., and Peng, Y. (2014). Recognition of similar objects using simulated prosthetic vision. *Artificial organs*, 38(2):159–167.

[Jones and Marc, 2005] Jones, B. W. and Marc, R. E. (2005). Retinal remodeling during retinal degeneration. *Experimental eye research*, 81(2):123–137.

[Kassim et al., 2016] Kassim, A. M., Yasuno, T., Suzuki, H., Aras, M. S. M., Jaafar, H. I., Jafar, F. A., and Subramonian, S. (2016). Conceptual design and implementation of electronic spectacle based obstacle detection for visually impaired persons. *Journal of Advanced Mechanical Design, Systems, and Manufacturing*, 10(7):JAMDSM0094–JAMDSM0094.

[Kelly et al., 2011] Kelly, S. K., Shire, D. B., Chen, J., Doyle, P., Gingerich, M. D., Cogan, S. F., Drohan, W. A., Behan, S., Theogarajan, L., Wyatt, J. L., et al. (2011). A hermetic wireless subretinal neurostimulator for vision prostheses. *IEEE transactions on biomedical engineering*, 58(11):3197–3205.

[Kendall et al., 2015] Kendall, A., Badrinarayanan, V., and Cipolla, R. (2015). Bayesian segnet : Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding. *arXiv preprint arXiv:1511.02680*.

[Keserü et al., 2012] Keserü, M., Feucht, M., Bornfeld, N., Laube, T., Walter, P., Rössler, G., Velikay-Parel, M., Hornig, R., and Richard, G. (2012). Acute electrical stimulation of the human retina with an epiretinal electrode array. *Acta ophthalmologica*, 90(1):e1–e8.

[Kim et al., 2009] Kim, L., Park, S., Lee, S., and Ha, S. (2009). An electronic traveler aid for the blind using multiple range sensors. *IEICE Electronics Express*, 6(11):794–799.

[Kitiratschky et al., 2015] Kitiratschky, V. B., Stingl, K., Wilhelm, B., Peters, T., Besch, D., Sachs, H., Gekeler, F., Bartz-Schmidt, K. U., and Zrenner, E. (2015). Safety evaluation of "retina implant alpha ims"—a prospective clinical trial. *Graefe's Archive for Clinical and Experimental Ophthalmology*, 253(3):381–387.

[Kotecha et al., 2014] Kotecha, A., Zhong, J., Stewart, D., and da Cruz, L. (2014). The argus ii prosthesis facilitates reaching and grasping tasks : a case series. *BMC ophthalmology*, 14(1):1–8.

[Krizhevsky et al., 2012] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.

[Kumar et al., 2014] Kumar, K., Champaty, B., Uvanesh, K., Chachan, R., Pal, K., and Anis, A. (2014). Development of an ultrasonic cane as a navigation aid for the blind people. In *2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)*, pages 475–479. IEEE.

[LeCun et al., 1998] LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.

[Li et al., 2018] Li, H., Su, X., Wang, J., Kan, H., Han, T., Zeng, Y., and Chai, X. (2018). Image processing strategies based on saliency segmentation for object recognition under simulated prosthetic vision. *Artificial intelligence in medicine*, 84:64–78.

[Lieby et al., 2011] Lieby, P., Barnes, N., McCarthy, C., Liu, N., Dennett, H., Walker, J. G., Botea, V., and Scott, A. F. (2011). Substituting depth for intensity and real-time phosphene rendering : Visual navigation under low vision conditions. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 8017–8020. IEEE.

[Long et al., 2015] Long, J., Shelhamer, E., and Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440.

[Lowery et al., 2015] Lowery, A. J., Rosenfeld, J. V., Lewis, P. M., Browne, D., Mohan, A., Brunton, E., Yan, E., Maller, J., Mann, C., Rajan, R., et al. (2015). Restoration of vision using wireless cortical implants : The monash vision group project. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1041–1044. IEEE.

[Lowery et al., 2017] Lowery, A. J., Rosenfeld, J. V., Rosa, M. G., Brunton, E., Rajan, R., Mann, C., Armstrong, M., Mohan, A., Josh, H., Kleeman, L., et al. (2017). Monash vision group's gennaris cortical implant for vision restoration. In *Artificial Vision*, pages 215–225. Springer.

[Lu et al., 2013] Lu, Y., Kan, H., Liu, J., Wang, J., Tao, C., Chen, Y., Ren, Q., Hu, J., and Chai, X. (2013). Optimizing chinese character displays improves recognition and reading performance of simulated irregular phosphene maps. *Investigative ophthalmology & visual science*, 54(4):2918–2926.

[Macé et al., 2015] Macé, M. J.-M., Guivarch, V., Denis, G., and Jouffrais, C. (2015). Simulated prosthetic vision : the benefits of computer-based object recognition and localization. *Artificial organs*, 39(7):E102–E113.

[Mahalle, 2014] Mahalle, S. (2014). Ultrasonic spectacles & waist-belt for visually impaired & blind person. *IOSR Journal of Engineering*, 4:46–49.

[Mallya and Lazebnik, 2015] Mallya, A. and Lazebnik, S. (2015). Learning informative edge maps for indoor scene layout prediction. In *Proceedings of the IEEE international conference on computer vision*, pages 936–944.

[Mandel et al., 2013] Mandel, Y., Goetz, G., Lavinsky, D., Huie, P., Mathieson, K., Wang, L., Kamins, T., Galambos, L., Manivanh, R., Harris, J., et al. (2013). Cortical responses elicited by photovoltaic subretinal prostheses exhibit similarities to visually evoked potentials. *Nature communications*, 4(1):1–9.

[Massé and Buhannic, 2017] Massé, A. and Buhannic, L. (2017). Comprendre la dégénérescence maculaire liée à l'âge. *Actualités Pharmaceutiques*, 56(565):18–21.

[Mathieson et al., 2012] Mathieson, K., Loudin, J., Goetz, G., Huie, P., Wang, L., Kamins, T. I., Galambos, L., Smith, R., Harris, J. S., Sher, A., et al. (2012). Photovoltaic retinal prosthesis with high pixel density. *Nature photonics*, 6(6):391–397.

[Maynard, 2001] Maynard, E. M. (2001). Visual prostheses. *Annual review of biomedical engineering*, 3(1):145–168.

[McCarthy et al., 2014] McCarthy, C., Walker, J. G., Lieby, P., Scott, A., and Barnes, N. (2014). Mobility and low contrast trip hazard avoidance using augmented depth. *Journal of neural engineering*, 12(1):016003.

[McCulloch and Pitts, 1943] McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133.

[Minge et al., 2017] Minge, M., Thüring, M., Wagner, I., and Kuhr, C. V. (2017). The mecue questionnaire : a modular tool for measuring user experience. In *Advances in Ergonomics Modeling, Usability & Special Populations*, pages 115–128. Springer.

[Minsky and Papert, 1969] Minsky, M. and Papert, S. (1969). An introduction to computational geometry. *Cambridge tiass., HIT*, 479:480.

[Mocanu et al., 2016] Mocanu, B., Tapu, R., and Zaharia, T. (2016). When ultrasonic sensors and computer vision join forces for efficient obstacle detection and recognition. *Sensors*, 16(11):1807.

[Moreno et al., 2012] Moreno, M., Shahrabadi, S., José, J., du Buf, J. H., and Rodrigues, J. M. (2012). Realtime local navigation for the blind : detection of lateral doors and sound interface. *Procedia Computer Science*, 14:74–82.

[Muqit et al., 2019] Muqit, M. M., Velikay-Parel, M., Weber, M., Dupeyron, G., Audemard, D., Corcostegui, B., Sahel, J., and Le Mer, Y. (2019). Six-month safety and efficacy of the intelligent retinal implant system ii device in retinitis pigmentosa. *Ophthalmology*, 126(4):637–639.

[Nanduri et al., 2008] Nanduri, D., Humayun, M., Greenberg, R., McMahon, M., and Weiland, J. (2008). Retinal prosthesis phosphene shape analysis. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1785–1788. IEEE.

[Niketeghad and Pouratian, 2019] Niketeghad, S. and Pouratian, N. (2019). Brain machine interfaces for vision restoration : the current state of cortical visual prosthetics. *Neurotherapeutics*, 16(1):134–143.

[Normann et al., 2009] Normann, R. A., Greger, B. A., House, P., Romero, S. F., Pelayo, F., and Fernandez, E. (2009). Toward the development of a cortically based visual neuroprosthesis. *Journal of neural engineering*, 6(3):035001.

[Palanker et al., 2020] Palanker, D., Le Mer, Y., Mohand-Said, S., Muqit, M., and Sahel, J. A. (2020). Photovoltaic restoration of central vision in atrophic age-related macular degeneration. *Ophthalmology*, 127(8):1097–1104.

[Panetsos et al., 2011] Panetsos, F., Sanchez-Jimenez, A., Cerio, E. D.-d., Diaz-Guemes, I., and Sanchez, F. M. (2011). Consistent phosphenes generated by electrical microstimulation of the visual thalamus. an experimental approach for thalamic visual neuroprostheses. *Frontiers in neuroscience*, 5:84.

[Parikh et al., 2013] Parikh, N., Itti, L., Humayun, M., and Weiland, J. (2013). Performance of visually guided tasks using simulated prosthetic vision and saliency-based cues. *Journal of neural engineering*, 10(2):026017.

[Pezaris and Eskandar, 2009] Pezaris, J. S. and Eskandar, E. N. (2009). Getting signals into the brain : visual prosthetics through thalamic microstimulation. *Neurosurgical focus*, 27(1):E6.

[Pissaloux et al., 2008] Pissaloux, E., Velázquez, R., and Maingreaud, F. (2008). Intelligent glasses : A multimodal interface for data communication to the visually impaired. In *2008 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pages 120–124. IEEE.

[Posch et al., 2014] Posch, C., Serrano-Gotarredona, T., Linares-Barranco, B., and Delbruck, T. (2014). Retinomorphic event-based vision sensors : bioinspired cameras with spiking output. *Proceedings of the IEEE*, 102(10):1470–1484.

[Ratelle, 2019] Ratelle, A. (2019). *Manuel d'intervention en orientation et mobilité.* Presses de l'Université de Montréal PUM.

[Rizzo III, 2011] Rizzo III, J. F. (2011). Update on retinal prosthetic research : the boston retinal implant project. *Journal of Neuro-ophthalmology*, 31(2):160–168.

[Rodieck, 2003] Rodieck, R. W. (2003). *La vision.* De Boeck Supérieur.

[Ronneberger et al., 2015] Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net : Convolutional networks for biomedical image segmentation. In *International Confe-*

rence on Medical image computing and computer-assisted intervention, pages 234–241. Springer.

[Rosenblatt, 1958] Rosenblatt, F. (1958). The perceptron : a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386.

[Sadi et al., 2014] Sadi, M. S., Mahmud, S., Kamal, M. M., and Bayazid, A. I. (2014). Automated walk-in assistant for the blinds. In *2014 International Conference on Electrical Engineering and Information & Communication Technology*, pages 1–4. IEEE.

[Sanchez-Garcia et al., 2018] Sanchez-Garcia, M., Martinez-Cantin, R., and Guerrero, J. J. (2018). Structural and object detection for phosphene images. *arXiv preprint arXiv:1809.09607*.

[Sanchez-Garcia et al., 2019] Sanchez-Garcia, M., Martinez-Cantin, R., and Guerrero, J. J. (2019). Indoor scenes understanding for visual prosthesis with fully convolutional networks. In *Visigrapp (5 : Visapp)*, pages 218–225.

[Sanchez-Garcia et al., 2020] Sanchez-Garcia, M., Martinez-Cantin, R., and Guerrero, J. J. (2020). Semantic and structural image segmentation for prosthetic vision. *Plos one*, 15(1):e0227677.

[Scherlen et al., 2007] Scherlen, A.-C., Dumas, J. C., Guedj, B., and Vignot, A. (2007). " recognizecane" : The new concept of a cane which recognizes the most common objects and safety clues. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6356–6359. IEEE.

[Schmidt et al., 1996] Schmidt, E. M., Bak, M. J., Hambrecht, F., Kufta, C. V., O'rourke, D., and Vallabhanath, P. (1996). Feasibility of a visual prosthesis for the blind based on intracortical micro stimulation of the visual cortex. *Brain*, 119(2):507–522.

[Sharma et al., 2017] Sharma, S., Gupta, M., Kumar, A., Tripathi, M., and Gaur, M. S. (2017). Multiple distance sensors based smart stick for visually impaired people. In *2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC)*, pages 1–5. IEEE.

[Shim and Yoon, 2002] Shim, I. and Yoon, J. (2002). A robotic cane based on interactive technology. In *IEEE 2002 28th Annual Conference of the Industrial Electronics Society. IECON 02*, volume 3, pages 2249–2254. IEEE.

[Simonyan and Zisserman, 2014] Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

[Sinclair et al., 2016] Sinclair, N. C., Shivdasani, M. N., Perera, T., Gillespie, L. N., McDermott, H. J., Ayton, L. N., Blamey, P. J., Consortium, B. V. A., et al. (2016). The appearance of phosphenes elicited using a suprachoroidal retinal prosthesis. *Investigative ophthalmology & visual science*, 57(11):4948–4961.

[Srinivas et al., 2020] Srinivas, K. S., Sahithya, K., Tejaswi, G. L., Gopal, K. H., and Karthik, B. P. (2020). A new method for recognition and obstacle detection for visually challenged using smart glasses powered with raspberry pi. *International Journal of Engineering Applied Sciences and Technology*, 5(1):408–412.

[Stingl et al., 2013] Stingl, K., Bartz-Schmidt, K. U., Besch, D., Braun, A., Bruckmann, A., Gekeler, F., Greppmaier, U., Hipp, S., Hörtdörfer, G., Kernstock, C., et al. (2013). Artificial vision with wirelessly powered subretinal electronic implant alpha-ims. *Proceedings of the Royal Society B : Biological Sciences*, 280(1757):20130077.

[Stingl et al., 2015] Stingl, K., Bartz-Schmidt, K. U., Besch, D., Chee, C. K., Cottriall, C. L., Gekeler, F., Groppe, M., Jackson, T. L., MacLaren, R. E., Koitschev, A., et al. (2015). Subretinal visual implant alpha ims - clinical trial interim report. *Vision research*, 111:149–160.

[Stoll et al., 2015] Stoll, C., Palluel-Germain, R., Fristot, V., Pellerin, D., Alleysson, D., and Graff, C. (2015). Navigating from a depth image converted into sound. *Applied bionics and biomechanics*, 2015.

[Suaning et al., 2014] Suaning, G. J., Lovell, N. H., and Lehmann, T. (2014). Neuromodulation of the retina from the suprachoroidal space : the phoenix 99 implant. In *2014 IEEE Biomedical Circuits and Systems Conference (BioCAS) Proceedings*, pages 256–259. IEEE.

[Tan and Le, 2019] Tan, M. and Le, Q. (2019). Efficientnet : Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR.

[Tatur, 2011] Tatur, G. (2011). *Conception d'un système de vision par phosphènes*. PhD thesis, Montpellier 2.

[Tatur, 2018] Tatur, G. (2018). Scene representation for mobility of the visually impaired. *Mobility of Visually Impaired People : Fundamentals and ICT Assistive Technologies*, pages 283–310.

[Tusa et al., 1978] Tusa, R., Palmer, L., and Rosenquist, A. (1978). The retinotopic organization of area 17 (striate cortex) in the cat. *Journal of Comparative Neurology*, 177(2):213–235.

[van Steveninck et al., 2022] van Steveninck, J. d. R., van Gestel, T., Koenders, P., van der Ham, G., Vereecken, F., Güçlü, U., van Gerven, M., Güçlütürk, Y., and van Wezel, R. (2022). Real-world indoor mobility with simulated prosthetic vision : The benefits and feasibility of contour-based scene simplification at different phosphene resolutions. *Journal of Vision*, 22(2):1–1.

[Velázquez, 2017] Velázquez, R. (2017). Mobility of visually impaired people : fundamentals and ict assistive technologies. *OPENAIRE*.

[Veraart et al., 1998] Veraart, C., Raftopoulos, C., Mortimer, J. T., Delbeke, J., Pins, D., Michaux, G., Vanlierde, A., Parrini, S., and Wanet-Defalque, M.-C. (1998). Visual sensations produced by optic nerve stimulation using an implanted self-sizing spiral cuff electrode. *Brain research*, 813(1):181–186.

[Vergnieux et al., 2014] Vergnieux, V., Macé, M. J.-M., and Jouffrais, C. (2014). Wayfinding with simulated prosthetic vision : Performance comparison with regular and structure-enhanced renderings. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2585–2588. IEEE.

[Vergnieux et al., 2017] Vergnieux, V., Macé, M. J.-M., and Jouffrais, C. (2017). Simplification of visual rendering in simulated prosthetic vision facilitates navigation. *Artificial organs*, 41(9):852–861.

[Wang et al., 2013a] Wang, C., Brunton, E., Haghgooie, S., Cassells, K., Lowery, A., and Rajan, R. (2013a). Characteristics of electrode impedance and stimulation efficacy of a chronic cortical implant using novel annulus electrodes in rat motor cortex. *Journal of neural engineering*, 10(4):046010.

[Wang et al., 2008] Wang, L., Yang, L., and Dagnelie, G. (2008). Virtual wayfinding using simulated prosthetic vision in gaze-locked viewing. *Optometry and vision science : official publication of the American Academy of Optometry*, 85(11):E1057.

[Wang et al., 2013b] Wang, S., Yang, X., and Tian, Y. (2013b). Detecting signage and doors for blind navigation and wayfinding. *Network Modeling Analysis in Health Informatics and Bioinformatics*, 2(2):81–93.

[Wankhade et al., 2017] Wankhade, S., Bichukale, M., Desai, S., Kamthe, S., and Borate, A. (2017). Smart stick for blind people with live video feed. *International Research Journal of Engineering and Technology*, 4(3):1774–1778.

[Wilson and Dorman, 2008a] Wilson, B. S. and Dorman, M. F. (2008a). Cochlear implants : a remarkable past and a brilliant future. *Hearing research*, 242(1-2):3–21.

[Wilson and Dorman, 2008b] Wilson, B. S. and Dorman, M. F. (2008b). Cochlear implants : current designs and future possibilities. *J Rehabil Res Dev*, 45(5):695–730.

[Yi and Tian, 2015] Yi, C. and Tian, Y. (2015). Assistive text reading from natural scene for blind persons. In *Mobile Cloud Visual Media Computing*, pages 219–241. Springer.

[Zapf et al., 2014] Zapf, M. P., Matteucci, P. B., Lovell, N. H., Zheng, S., and Suaning, G. J. (2014). Towards photorealistic and immersive virtual-reality environments for simulated prosthetic vision : Integrating recent breakthroughs in consumer hardware and software. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2597–2600. IEEE.

[Zapf et al., 2015] Zapf, M. P. H., Boon, M.-Y., Matteucci, P. B., Lovell, N. H., and Suaning, G. J. (2015). Towards an assistive peripheral visual prosthesis for long-term treat-

ment of retinitis pigmentosa : evaluating mobility performance in immersive simulations. *Journal of neural engineering*, 12(3):036001.

[Zhao et al., 2010] Zhao, Y., Lu, Y., Tian, Y., Li, L., Ren, Q., and Chai, X. (2010). Image processing based recognition of images with a limited number of pixels using simulated prosthetic vision. *Information Sciences*, 180(16):2915–2924.

[Zou et al., 2018] Zou, C., Colburn, A., Shan, Q., and Hoiem, D. (2018). Layoutnet : Reconstructing the 3d room layout from a single rgb image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2051–2059.

[Zrenner et al., 2011] Zrenner, E., Bartz-Schmidt, K. U., Benav, H., Besch, D., Bruckmann, A., Gabel, V.-P., Gekeler, F., Greppmaier, U., Harscher, A., Kibbel, S., et al. (2011). Subretinal electronic chips allow blind patients to read letters and combine them to words. *Proceedings of the Royal Society B : Biological Sciences*, 278(1711):1489–1497.

[Zrenner et al., 2017] Zrenner, E., Bartz-Schmidt, K. U., Besch, D., Gekeler, F., Koitschev, A., Sachs, H. G., and Stingl, K. (2017). The subretinal implant alpha : implantation and functional results. In *Artificial Vision*, pages 65–83. Springer.

[Zrenner et al., 2009] Zrenner, E., Wilke, R., Sachs, H., Bartz-Schmidt, K., Gekeler, F., Besch, D., Benav, H., Bruckmann, A., Greppmaier, U., Harscher, A., et al. (2009). Subretinal microelectrode arrays implanted into blind retinitis pigmentosa patients allow recognition of letters and direction of thin stripes. In *World Congress on Medical Physics and Biomedical Engineering, September 7-12, 2009, Munich, Germany*, pages 444–447. Springer.

Popularized abstract

In the field of assistance to the blind, many devices have been developed. In particular, visual neuroprostheses are currently undergoing significant development, which has led to the marketing of several of them. These neuroprostheses can be implanted in different areas of the human visual system and can generate the appearance of luminous spots in the field of vision of blind people, called phosphenes.

The objective of our work is to provide an additional tool, associated with a visual neuroprosthesis, which exploits a multimodal (addition of audio clues for example) or interactive aspect. For this, we have conducted two experiments that evaluate the impact of of the addition of an audio system and an interactive system on the performances of subjects in different different contexts.

We detail in this manuscript the rendering construction methods we use, present the developed prosthetic vision simulator, as well as the results we obtained during our two experiments.

Résumé vulgarisé

Dans le domaine de l'assistance aux non-voyants, de nombreux dispositifs ont vu le jour. Plus particulièrement, les neuroprothèses visuelles connaissent à l'heure actuelle un développement conséquent, qui a conduit à la commercialisation de plusieurs d'entre elles. Ces neuroprothèses peuvent être implantées à différents endroits du système visuel humain et peuvent générer l'apparition de taches lumineuses dans le champ de vision des non-voyants, appelées phosphènes.

L'objectif de nos travaux est de fournir un outil supplémentaire, associé avec une neuroprothèse visuelle, qui exploite un caractère multimodal (ajout d'audio par exemple) ou interactif. Pour cela, nous avons conduit deux expériences qui évaluent l'impact de l'ajout d'un système sonore et d'un système interactif sur les performances de sujets dans différents contextes.

Nous détaillons dans ce manuscrit les méthodes de construction de rendu que nous utilisons, présentons le simulateur de vision prothétique développé, ainsi que les résultats que nous avons obtenus pendant nos deux expériences.

Abstract

According to the WHO, 253 million people suffer from visual impairment : 36 million of them are blind and 217 million have a visual impairment between moderate and severe. In France, there are 1.7 million visually impaired people, including 207,000 blind and profoundly visually impaired people and about 900,000 moderately visually impaired people. People with visual impairments traditionally use a cane or a guide dog to navigate. But canes and guide dogs have a limited role in understanding the visual scene.

Visual neuroprostheses are devices that appeared in the 1960s, which allow the restoration of limited visual perception in blind implanted patients. Some of these neuroprostheses, localized in the retina or visual cortex, include an implant, a computing device and an external camera to capture the scene. An impoverished visual perception is then restored when the retina or visual cortex is microstimulated via the implant, leading to the appearance of white dots called phosphenes. Over the last ten years, several implants have been placed on blind people and are in clinical trials.

The difficulty of access to blind people (whether implanted or not) constraints us to use simulators. Their objective is to reproduce the real conditions of perception of phosphenes in order to get as close as possible to reality.

However, the resolution of the current implants (i.e. the number of electrodes and their spacing) that have passed the clinical trial phases remains low. This low resolution, associated with the limited number of different colors rendered by the implants, limits the information that can be transferred and thus perceived.

In this document we present the different technologies developed over the last 30 years with the objective of allowing blind people with implants to perceive their environment by generating coherent phosphenic maps. We have set up two comparative studies between different renderings and we have shown, on the one hand, that the addition of sound cues coupled with a visual rendering allows the subjects to better understand the scene and to navigate in it, and on the other hand, that giving the blind the possibility to alternate between several rendering modes in real time significantly increases their understanding of the environment.

Résumé

Le dernier rapport de l'OMS sur la non-voyance rapporte que 253 millions de personnes souffrent de déficience visuelle (36 millions d'entre elles sont aveugles et 217 millions ont une déficience visuelle). Pour aider les non-voyants dans leurs tâches quotidiennes, de nombreux dispositifs ont été conçus.

Les neuroprothèses visuelles sont des dispositifs apparus dans les années 60, qui permettent de rétablir une perception visuelle limitée chez les patients non-voyants implantés. Certaines de ces neuroprothèses, implantées dans la rétine ou dans le cortex visuel, comprennent un implant, un dispositif de calcul informatique et une caméra externe pour capturer la scène. Une perception visuelle appauvrie sous forme d'un ensemble de points blancs appelés phosphènes est alors restaurée grâce à une micro-stimulation de la rétine ou du cortex visuel via l'implant. Pendant les dix dernières années, plusieurs dispositifs ont été implantés sur des non-voyants et sont en phase d'essai clinique. Toutefois, la résolution (c'est-à-dire le nombre d'électrodes et leur densité) des implants actuels reste faible. Cette faible résolution, associée au nombre limité de couleurs différentes rendues par les implants, limite les informations qui peuvent être transférées et donc perçues.

Au cours de nos travaux, nous nous sommes intéressés à l'intérêt et à la manière de coupler des mécanismes interactifs et multimodaux avec une neuroprothèse visuelle, dans l'objectif d'améliorer les capacités de décisions des non-voyants. Il s'agit plus précisément de savoir dans quelles mesures l'ajout d'un système d'indices audio permet aux non-voyants de se déplacer plus facilement, et dans quelles mesures l'ajout d'un système de sélection interactif de rendus permet de désambiguïser la scène.

Dans ce mémoire, nous présentons les différentes technologies développées ces 30 dernières années ayant pour objectif de permettre à des non-voyants implantés de percevoir leur environnement en générant des cartes phosphéniques cohérentes. Nous abordons ensuite la question de la simulation de la vision prothétique, principalement due à la difficulté d'accès aux patients implantés. Ces simulateurs ont pour objectif de reproduire les conditions réelles de perception de phosphènes pour s'approcher le plus possible de la réalité. Ils permettent également de tester des méthodes de rendu originales. Nous présentons enfin deux études comparatives entre différents rendus, et nous montrons, d'une part, que l'ajout d'indices sonores couplés avec une restitution visuelle permet aux sujets de mieux comprendre la scène et de naviguer dans cette dernière, et, d'autre part, que donner aux non-voyants la possibilité d'alterner entre plusieurs modes de rendu en temps réel augmente significativement leur compréhension de l'environnement. En ce qui concerne l'ajout d'indices sonores, nous montrons que celui-ci permet aux sujets

d'améliorer leurs capacités de perception de la scène (détection et localisation d'obstacles), et procure aux sujets un sentiment de confiance. En ce qui concerne le système interactif, nous montrons qu'il permet aux sujets d'avoir un meilleur taux de bonnes réponses à une série de questions réparties sur trois catégories (objets, rues, portes et passages piétons). De plus, les sujets montrent un intérêt certain pour ce système interactif.